The Dissertation Committee for Wei Chen certifies that this is the approved version of the following dissertation:

Multistage Stochastic Programming Models for the Portfolio Optimization of Oil Projects

Committee:

Leon S. Lasdon, Supervisor

James S. Dyer, Co-Supervisor

Anantaram Balakrishnan

Larry W. Lake

Christopher J. Jablonowski

Multistage Stochastic Programming Models for the Portfolio Optimization of Oil Projects

by

Wei Chen, B.E., M.MGMT, M.S.IROM

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas at Austin

August 2011

Dedication

To my dear parents, Bowen Chen and Guanzhen Zhu

and

my beloved wife, Shuming Liu.

Acknowledgements

I would like to express my deepest gratitude to my advisors Dr. Lasdon and Dr. Dyer for their invaluable encouragement, guidance, and inspiration from the very early stage to the final phase. This dissertation would not have been possible without their supervision. I am indebted a lot to Dr. Lasdon for his innumerous help and persistent support. My appreciation to him is beyond what words can express. I am grateful to Dr. Lake and Dr. Jablonowski for their insightful comments and suggestions. Their inputs have helped me to better align the dissertation to the oil industry's needs. I would like to thank Dr. Balakrishnan for serving in my committee. His seminar classes have enlightened many Ph.D. students including me.

It is a great pleasure to convey my gratitude to Dr. Yu, Dr. Bard, Dr. Morton, and Dr. Tompaidis for their guidance during my studies. My thanks also go to Dr. Bickel, Mr. Faya, and Dr. Brandao who have helped me to build up E&P models. Particularly, my wife, Dr. Liu, has been continuously providing insightful comments to my work on option pricing. Without her lasting encouragement, I might have missed enjoying the beauty of option pricing. I deeply appreciate my Master supervisor, Dr. Lan, in Tsinghua University in China, who introduced me to the fantastic area of optimization.

I appreciate the support from the faculty and staff of the Department of Information, Risk, and Operations Management at the University of Texas at Austin under many different circumstances. I gratefully acknowledge the financial support from the Center for Petroleum Asset Risk Management (CPARM) at UT Austin.

Last, but the most important, I would specially like to thank my wife again and our parents for their unconditional love, endless consideration, and constant encouragement.

Multistage Stochastic Programming Models for the Portfolio Optimization of Oil Projects

Wei Chen, Ph.D.

The University of Texas at Austin, 2011

Supervisors: Leon S. Lasdon

James S. Dyer

Exploration and production (E&P) involves the upstream activities from looking for promising reservoirs to extracting oil and selling it to downstream companies. E&P is the most profitable business in the oil industry. However, it is also the most capital-intensive and risky. Hence, the proper assessment of E&P projects with effective management of uncertainties is crucial to the success of any upstream business.

This dissertation is concentrated on developing portfolio optimization models to manage E&P projects. The idea is not new, but it has been mostly restricted to the conceptual level due to the inherent complications to capture interactions among projects. We disentangle the complications by modeling the project portfolio optimization problem as multistage stochastic programs with mixed integer programming (MIP) techniques.

Due to the disparate nature of uncertainties, we separately consider explored and unexplored oil fields. We model portfolios of real options and portfolios of decision trees for the two cases, respectively. The resulting project portfolio models provide rigorous and consistent treatments to optimally balance the total rewards and the overall risk.

For explored oil fields, oil price fluctuations dominate the geologic risk. The field development process hence can be modeled and assessed as sequentially compounded options with our optimization based option pricing models. We can further model the portfolio of real options to solve the dynamic capital budgeting problem for oil projects.

For unexplored oil fields, the geologic risk plays the dominating role to determine how a field is optimally explored and developed. We can model the E&P process as a decision tree in the form of an optimization model with MIP techniques. By applying the inventory-style budget constraints, we can pool multiple project-specific decision trees to get the multistage E&P project portfolio optimization (MEPPO) model. The resulting large scale MILP is efficiently solved by a decomposition-based primal heuristic algorithm.

The MEPPO model requires a scenario tree to approximate the stochastic process of the geologic parameters. We apply statistical learning, Monte Carlo simulation, and scenario reduction methods to generate the scenario tree, in which prior beliefs can be progressively refined with new information.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

Petroleum is the life-blood of modern industries and economies. According to studies performed by the Energy Information Administration (EIA) as of 2006-2009, petroleum accounts for 37% of all energy consumption in the United States. Specifically, petroleum makes up 84% of all the energy used in transportation and 30% in industry and manufacturing. The world oil demand has constantly grown at annualized rates fluctuating around 1.7% over the past four decades, which is consistent to the growth of the world GDP. However, most production wells in the world are aging after decades of producing and the production has started to decline. In the United States, the total production has been steadily declining after it reached the peak around 1970. More than half of the present production wells have been working for more than 15 years. Although alternative energy has attracted increasing attentions and is expected to grow constantly, both government research institutions like EIA and giant oil companies like BP believe that it is unlikely that alternative energy can substitute petroleum in any significant fraction in the foreseeable future prior to 2030. Therefore, to meet the demand growth and achieve success, major oil producers (whether nations or companies) have set aggressive goals and plans for reserve replacement and production growth. The goals can be reached by three fundamental approaches, exploration and production (E&P), merger and/or acquisition of proven prospects, and enhanced recovery of older fields. This dissertation considers the first two types of oil projects. We develop different optimization methods to model and assess them in a portfolio perspective.

Oil industry is broken down into three segments, upstream, midstream, and downstream. Exploration and production (E&P) involves the upstream activities from looking for promising reservoirs to extracting oil from underground reservoir rocks and selling it to downstream companies. The downstream companies are responsible to refine crude oil and market final products. The midstream business involves storage and transportation, just like a bridge to connect the other two. E&P represents the most profitable but also the most capital-intensive and risky business in the petroleum industry. The E&P investment decisions made today determine activities over the next few decades and have far-reaching impacts to future cash-flows, drilling plans, and production growth. For pure E&P companies and most integrated oil companies whose business covers all three segments, E&P activities are the major sources of profits and losses. Ball and Savage (1999) point out that "most exploration projects are total failures while a few are tremendously successful" and "a major (successful) discovery every decade or so can sustain a large company." Therefore, how to make good decisions to simultaneously secure production growth and effectively manage E&P risks is crucial to the success of these oil companies.

The dissertation develops quantitative models to assist the investors to select promising E&P projects and make optimal investment decisions by looking for opportunities embedded in all projects and then trading profits against uncertainties. This chapter serves as a brief introduction to the background of our approaches. In §1.1, we introduce the background and motivations of the problem we want to address. In §1.2, we present the research objectives to achieve through the dissertation. In §1.3, we review the main results and contributions of the dissertation. Finally, in §1.4, the organization of this dissertation is presented as a quick guide to read.

1.1 BACKGROUND AND MOTIVATIONS

E&P projects are risky investments that are characterized by intensive capital expenditures, sequential decisions, and complex uncertainties. Many oil companies have long used financial theories and quantitative methods to screen projects and make investment decisions. Both academic researchers and industrial practitioners have used optimization methods, decision analysis tools, and real options to assess and rank E&P projects. However, most practices are focused on individual project evaluation and static capital budgeting without recourse decisions.

There are three main methods to evaluate and compare investment projects: discounted cash flow (DCF) method, decision tree analysis, and real options. Each has different advantages and disadvantages, depending on application contexts. Copeland and Antikarov (2003) and Smith and Nau (1995) provide extensive discussions and comparisons on three methods and their proper application conditions. In summary, all methods aim to get a (expected) present values for alternative projects. The DCF method itself can not handle cases where cash flows depend on decisions. The real options method relies on the complete market assumption. The decision tree analysis, which is rooted at the stochastic dynamic programming, provides a unified approach to evaluate general risky projects. We can treat the lattice-based option pricing models as special decision trees.

Over the past two decades, more interests are centered on the portfolio approach to manage and develop risky projects. Our dissertation is one of such efforts to look the multiple E&P projects investment from a dynamic portfolio view. The portfolio view coincides with the industry's practice to spread geological and political risks over projects selected from various regions of different geological profiles. In the rest of this section, we conduct a literature review on the evaluation and selection of E&P projects.

The literatures cover topics like single project evaluation, capital budgeting problem, oil field project modeling, decision tree modeling, real options, project portfolio optimization, project dependence, risk measures, statistical learning.

1.1.1 Classical project valuation and selection methods

The concept of expected net present value (ENPV) is probably the single most important criterion to assess and compare projects. For deterministic projects whose cash flows are fixed, we can apply the bond valuation methods to evaluate them. The net present value (NPV) of a project equals to the difference between the sum of discounted future cash flows and the initial investment cost. The discount factor is the risk-free rate. For risky projects whose cash flows are random, the ENPV of a project is the expected NPVs over all possible scenarios. For each scenario, a different stream of cash flows (called a sample path) can be simulated and thus a different NPV can be computed. In this case, the discount factors should include proper risk premiums. For the more practical cases, the cash flows depend on the investor's decisions and risk attitude and possibly various constraints. Therefore, optimization models such as decision trees should be used to find optimal decisions and corresponding cash flows. Otherwise, the ENPV may not correctly reflect the project value. In theory, stochastic dynamic programming (SDP) provides a unified approach to model the general project valuation problem. In practice, however, it is very difficult to solve SDPs of any realistic size numerically, due to the curse of dimensionality (de Farias and Roy, 2003). In special cases like in this dissertation, we formulate, or approximate to be more accurate, the SDP with an equivalent deterministic linear program.

The simplest rule to screen projects is to compute their ENPVs one by one, drop those with negative ENPVs, and accept those with positive ENPVs and prefer projects with higher ENPVs. The major advantage of ENPV rule is that it captures the time value of money by appropriate discounting. By adding proper risk premium terms to the discount factor, we may take into account the impact of event-driven uncertainties to the project value in the way that we handle default risks in financial markets. Such events could be political events, geological disasters, or industrial disasters. If the budget is limited, not all positive-ENPV projects may be selected. We need solve a capital budgeting problem, which is essentially a knapsack problem, to allocate funds to maximize the total profits. When the number of potential projects is large, we can avoid solving the hard knapsack problem by applying simple heuristics algorithms to allocate the budget. One simple rule is to sort projects in the descent order of ENPVs and then allocate fund to the sorted projects in the same order until the budget is exhausted. A better and similar rule is to allocate fund to the projects in the descent order of their profitability indexes (PIs). Luenberger (1998), Brealey and Myers (2003), and Copeland and Antikarov (2003) are excellent textbooks providing detailed discussions on investment project valuation and selection.

1.1.2 Model individual exploration and production projects

In early ages when geological information was limited, people mainly paid attentions to create deterministic models to evaluate proven prospects. Those models consider the problems such as platform design (size, type, and location), well drilling, and reservoir production scheduling. The objectives are either to minimize development costs or to maximize overall benefits subject to physical, technical, and market restrictions. Usually, if the decisions are related to locations and schedules, 0-1 binary variables will be used.

Devine and Lesso (1972) consider the problem where to locate platforms and how to assign production wells to platforms. They formulated the problem as a two-dimensional facility location problem to minimize total drilling costs. Babayev (1975) model the well drilling problem on multilayer oil and gas fields to find the optimal number of wells to be drilled on each layer and to be transferred between layers such that the overall costs are minimal. Frair and Devine (1975) develop a model to optimally develop offshore oil and gas reserves in a two-phase style, first to design and locate platforms and then to schedule reservoir productions. The objective of the model is to maximize the after-tax profits subject to geological and technical constraints. Lasdon et al. (1986) study the problem of determining a production profile from a gas field to satisfy a pre-estimated demand profile. Haugland et al. (1988) extend previous work to consider problems to maximize the present value by finding optimal platform locations and capacities, drilling plans, and individual well production profiles.

The development of technology and data collection and analysis allowed people to use distributions to describe random parameters, which made quantitative methods applicable to take care of E&P uncertainties. Grayson (1960) and Kaufman (1963) are the earliest attempts to apply decision analysis to E&P projects. McCray (1975), Newendorp and Schuyler (1975, 2000), and Megil (1977) popularize the applications of decision analysis in oil industry. Since then, more robust and realistic quantitative models and methods have been created to select risky E&P projects.

Jornsten (1992) and Jonsbraten (1998) develop scenario-based optimization models to find optimal sequence to develop a number of oil fields. They consider various constraints such as multi-period budget constraints, required supplies, and production limits (due to storage and transportation capacities). The main differences between them lie in where the uncertainties come from and how to generate production profiles. Jornsten only considers uncertain demand, while Jonsbraten only considers price fluctuations. Both models are formulated as scenario-based MIPs and are solved with the progressive hedging algorithm (PHA), which is proposed by Rockafellar and Wets (1991). The idea is to decompose the stochastic program into individual scenario subproblems. Each scenario subproblem is solved as a deterministic optimization problem with one realization (sample path) of random parameters. Then a scenario aggregation technique is applied to the solutions of scenario subproblems to recover the so-called non-anticipativity solution. The aggregation is the reverse process of decomposing the scenario tree into parallel sample paths. It recovers the tree structure solution from solutions of separate problems. The original PHA method is designed to aggregate continuous decision variables. However, under certain cases, the method can be extended to aggregate integer or binary decisions with certain heuristic rules. Jornsten and Jonsbraten explain how the PHA method is applicable to their models.

Due to the similarities between traditional R&D projects and the E&P projects, such as "go-no go" type decisions and project precedence and interdependence, we may borrow many ideas from the literature of R&D project selection and development to model and assess E&P projects. Ghasemzadeh et al. (1999) develop a deterministic 0-1 model to solve dynamic capital budgeting and project scheduling problem. The model pays special attentions to capture various project timing and interdependences. Heidenberger (1996) proposes a seminal approach to model R&D projects as decision trees with MIP skills. The scenarios and decisions are modeled as chance nodes and decision nodes, respectively. The nodes are organized in a tree structure. One interesting property is that the transition probabilities of chance nodes may depend on decisions. However, the decision-dependent chance nodes may result quick growth in problem size. Heidenberger's work is among the earliest efforts to model project-specific decision trees as MIPs. After assuming all project-specific decision trees to share the same structure, the author further develops a dynamic budgeting problem such that the aggregated capital expenditures of all projects along any path are bounded from above by a given budget. As we will mention soon, Gustafsson and Salo (2005) extend Heidenberger's work to develop a more general project portfolio optimization model.

Heidenberger's MILP decision tree model can deal with various uncertainties if the underlying random parameters can be approximated by discrete distributions. When uncertainties are dominated by market risks, we may use the contingent claim analysis to valuate projects whose outputs are tradable commodities. The resulting methods are real options, the options (to make investment decisions) in real projects (as opposed to financial securities). A real option is the right, but not the obligation, to take an action (e.g., deferring, expanding, contracting, or abandoning) at a predetermined cost called the exercise price, for a predetermined period of time-the life of the option (Copeland and Antikarov, 2003).

After decades of development, there are abundant publications on the theories and applications of real options. Dixit and Pindyck (1994), Trigeorgis (1996), and Copeland and Antikarov (2003) are three popular textbooks on real options. The former two are relatively more theoretical and the latter one is closer to the practitioner. All of them provide simple examples on the applications to evaluate oil projects. In general, oil projects are long-term projects and can be broken into a sequence of compound options, with one option's payoff being written on the value of successive options. Therefore, a sequential option pricing problem has to be solved to find the present value of the oil projects. Most oil projects' development durations are determined by lease terms. If certain activities are not finished upon required deadlines, the lease may be recalled and worthless. As a consequence, oil projects are modeled as sequential compound

American options (Dixit and Pindyck, 1994, Chapter 10). We delay more detailed discussions on general real options method for Chapter 3, where we create optimization models to solve the pricing problem of real options.

Smith and Mccardle (1999) consider a full life-time oil and gas project. To capture option values which have been missed by the DCF analysis, they allow the investor can wait and delay some investments and model such flexibilities as American options. They show how the real options approach is superior than traditional approach in which all investment decisions are made at the beginning and then all uncertainties are resolved and cash flow are determined. As Dixit and Pindyck (1994) point out, the real options approach allows a separate optimization in each of the contingencies of resolved uncertainties, whereas immediate action must be based on only the average scenario. This ability to tailor action to contingency gives value to the extra freedom to wait. Smith and Mccardle (1999) further propose a decision tree approach equivalent to the real options approach. To achieve the equivalence, proper discount factors must be used to reflect proper risk adjustments. The authors show how to find the proper discount factors with option pricing theories. Smith and Nau (1995) and Smith and Mccardle (1998) discuss the relationship between decision tree analysis and real options in details. They conclude that, if implemented correctly, both methods should give consistent results.

Lund (2001) presents one of the most comprehensive models to assess E&P projects. The model divides the life cycle of an E&P project into four phases, exploration, conceptual study, engineering and construction, and production. At each phase, the investor has flexibility (options) to decide whether to continue or not. If the project is not to continue at some point, it is abandoned henceforth. Lund considers four uncertain parameters: oil price, reservoir volume, well rate, and decline rate. Lund's paper provides an excellent example to understand the E&P business and we take much of the concepts to build our MILP decision tree models for E&P projects in Chapter 4.

1.1.3 Project portfolios and risk management

Since E&P projects are the most profitable and also the most risky venture in the whole petroleum industry, effective risk management is substantial to the success of the upstream business. Motivated by the Markowtiz's quantitative portfolio selection model, many scholars have made efforts to extend the mean-risk model from financial markets to quantify investment project uncertainties and select project portfolios.

To model the investor's preferences to the asymmetric distributions of project returns, Cozzolino (1974, 1977) applies expected (exponential) utility theory to model the optimization problem to select a portfolio of risky projects. The utility functions are chosen to reflect the investor's attitude towards rewards and risks. If the investor makes decisions rationally and consistently, then the risk behavior implied by the decisions can be described by the parameterized utility function. In Cozzolino's models, the objectives are to maximize the risk-adjusted profit of the portfolio. The risk-adjusted profit is essentially the certainty equivalent (CE) that is equal to the expected value less a risk premium term. Cozzolino further analytically discusses relevant topics such as risk attitude, risk sharing, and budget constraints.

Ball and Savage (1999) provide more extensive discussions on the differences of risk management between E&P projects and the financial markets. The authors claim that it is essential to take a holistic view to hedge local uncertainties, i.e., the project-specific geological risks, in a portfolio way. Particularly, they point out that a proper risk measure should be chosen to penalize larger losses more than small ones. As an example, they create a static scenario optimization model to determine the optimal mix of exploration and production projects and allocate funds to the selected projects. They use the expected downside risk as the risk measure. The downside risk can be the deviation below either the mean return or zero. The latter case essentially measures actual losses. The problem is to maximize overall return while minimizing expected losses. Similar to the traditional mean-variance portfolio model, the resulting exploration and production portfolio optimization (EPPO) model can be formulated in different ways to generate the efficient frontier of the portfolio. Each point on the efficient frontier corresponds to a mix of projects and the corresponding fund allocation decision.

In decision science society, people have extended the mean-risk model to the more general multi-attribute utility theory (MAUT, Keeney and Raiffa, 1976, Clemen and Reilly, 2004), which attempts to integrate various corporate objectives and risk policies into a unified measure to guide investment choices. Walls (1995) and Walls and Dyer (1996) employ the MAUT approach to model the conflicting decisions for large oil and gas companies.

Observing advantages and shortcomings of previous quantitative models to manage project risks, Gustafsson and Salo (2005) proposed an innovative contingent portfolio programming (CPP) framework for the management of risky projects. They use a state tree to capture the evolution of project uncertainties and model each project as a decision tree with MILP techniques. Those projects are allowed to interact by pooling their incurred cash flows together in the dynamic budget constraints. The CPP model allows a wide range of risk attitudes by using different utility functions. However, since the CPP model is a multi-stage stochastic optimization problem, usually a linear risk measure is chosen to ensure the resulting large scale problem is solvable. The CPP model and the EPPO model together serve as our starting point for the modeling of multistage E&P project portfolio model. We will delve into the modeling details in Chapter 4 for the optimization model and Chapter 5 for the scenario tree generation.

Meier et al. (2001) model project portfolios from the view of treating investment projects as real options when the outputs of the projects are tradable assets. So the uncertainties of the projects can be perfectly hedged away by trading in financial markets (Dixit and Pindyck, 1994). In Meier's model, all embedded real options are assumed to be American options with infinite expiry to take advantage of the closed-form optimal exercise strategies. This property restricts the application of their model even if the optimal exercise strategies can be numerically generated since the dependence on optimal exercise strategies rules out some investment opportunities which is not optimal for a specific project but may be optimal from a portfolio view.

Motivated by Longstaff and Schwartz (2001) and Meier et al (2001), we develop a general framework to price options and solve the dynamic capital budgeting problem with integer programming techniques. Our model is an extension to Meier's in that our numerical experiments show that the new dynamic capital budgeting model can capture investment opportunities which would have been missed by Meier's model.

1.1.4 Project dependence and statistical learning

When investment projects are managed in a portfolio, the interactions among them can be captured by (dynamic) budget constraints. Necessary side constraints may be applied as well to handle special interdependence relationships such as mutual exclusion, preemption, and precedence. However, besides the operational interdependence, statistical project dependence also plays an important role to affect the value of project portfolios.

The importance of assessing statistical dependence among multiple random variables has been well recognized by the decision science society. There are different ways to measure statistical dependence to different strengths. Three popular methods are (pairwise) correlations, marginal and conditional distributions, and (full or partial) joint distributions, listed in the ascent order of power. Clemen et al. (2000) also consider several other descriptive methods to describe and assess dependence. They review and compare different methods to assess dependence measures and present experimental results. They point out that a desirable dependence measure and assessment method should be rigorous in probability theory, consistent in various situations, and intuitive in interpretations. They suggest that the most accurate way to obtain a subjective dependence measure is to get expert options about pairwise correlations.

Although full joint distributions of multiple random variables are the most powerful way to describe statistical relationships, they are not applicable in general since normally they are not directly available. The typical way to recover joint distributions is to indirectly assess corresponding marginal and conditional distributions. However, this approach requires an exponentially growing number of conditional assessments and is only limited to small cases. Clemen and Reilly (1999) describe an alternative way to recover joint distribution by combining a specially chosen copula function and pairwise correlations. Makridakis and Winkler (1983) and Clemen and Winkler (1999) propose different methods to get improved estimations by "averaging" several different assessment methods or multiple experts' options.

When full joint distributions are not readily available, an approximation with ensured quality can be of great help as well. Smith (1993) extensively discusses the applications of a large class of methods, called moment methods, to construct discrete approximation to the continuous distributions. One application called moment matching can be formulated as a constrained optimization problem. In this problem, the decision variables are a discrete distribution measure, the constraints ensure that the moments of the chosen distribution measure match required moments, and the objective function is normally some distance metric defined on the probability space. The entropy method is a special moment matching problem whose objective function is related to the entropy function of the present distribution. Luenberger (1984), Miller and Liu (2002), and Bickel and Smith (2006) provide theories and examples to the entropy maximization methods.

In Chapter 5, we introduce two types of statistical dependence, inter-project and intra-project dependence, to characterize the joint distributions for our E&P project portfolio model. Based on the dependences, we incorporate two statistical learning to refine the investor's belief about the performance of considering reservoirs.

1.2 RESEARCH OBJECTIVES

Given a number of E&P projects to invest, the goal of our research is to develop quantitative models to assist the investor to determine the best mix of E&P projects and find the optimal contingent strategies to develop them. We require the strategies are taken by simultaneously considering opportunities available to all projects but not in isolation and project by project. We also want the optimality criteria to be aligned to the investor's reward-risk propensity. Consequently, the problem is to find the efficient project portfolio and dynamic portfolio strategies such that for each level of risk, the selected portfolio and strategies would return the maximum reward.

Since E&P projects are long-term risky investments, we need a probabilistic model to describe the evolution of uncertainties. Particularly, we are interested in proper ways to capture statistical dependences among projects and incorporate statistical learning so that the probabilistic model can be progressively refined with sequential observations.

We want to use the models to help the investor to rethink the investment opportunities in a systematic way, evaluate the marginal values of one or more projects when they are added to the existing portfolio, and foresee the impact of specific risk factors to the portfolio value. However, we should not rely on the solutions of the models to design exact development strategies and ask the investor to follow them mindlessly. Instead, the model and the solution is merely a

1.3 MAIN RESULTS AND CONTRIBUTIONS

Our research and contributions can be broken into three parts. First, we develop optimization-based option pricing framework to assess explored oil prospects. We then construct a dynamic capital budgeting model to manage oil projects. The model is convenient to implement and the outcomes convey important economic insights to understand the project assessment and prioritization under budget restrictions and market dynamics. Second, we model E&P projects as decision trees with MILP skills and implement project portfolio optimization to find optimal mix of projects. We analyze the full life-cycle of E&P projects and identify major uncertainties and managerial flexibilities. Finally, we identify statistical project dependences and incorporate learning effects into the process to generate scenario tree for the project portfolio model. We also apply scenario reduction schemes to reduce the scenario tree such that the resulting problem can be solved in reasonable time. Extensive numerical experiments are performed to verify the models and algorithms.

1.4 DISSERTATION ORGANIZATION

The dissertation is organized as follows. In Chapter 2, we give a brief review to the upstream business and discuss investment decisions in different phases during the life cycle of typical E&P projects. The resulting E&P phase model plays an important role in the development of our project portfolio model. Besides the upstream business review, we introduce a dynamic tank model which allows sequential drilling decisions to predict the oil field production. The dynamic tank model will be applied in Chapter 3 to model the dynamic capital budgeting problem for multiple oil projects. This chapter mainly serves as a background knowledge for modeling upstream activities.

In Chapter 3, we develop optimization based options pricing models and apply them to assess proven but undeveloped oil fields. Our ultimate goal is to determine the best project mix and the optimal drilling plans for the selected fields, under budget constraints and market uncertainty. To achieve this goal, we develop a dynamic capital budgeting model in which each project is modeled as compound sequential options using our optimization based option pricing framework. Extensive numerical experiments shows that our option pricing models and capital budgeting model have many advantages over traditional approaches, such as easy to implement and flexible to extend. Particularly, our model can capture investment opportunities which would have been missed by other approaches.

In Chapter 4, we consider the assessment and prioritization for unexplored oil fields, each of which is called an exploration and production (E&P) project. Unlike the projects considered in Chapter 3, E&P projects represent the most risky business in oil industry since the information about the field properties is rather rare. We take a "holistic" view (Ball and Savage, 1999) to evaluate E&P projects and manage incurred risks in portfolio level by taking all uncertainties and opportunities into account. We want to maximize the value of the portfolio while minimizing overall investment risks. Using a linear risk measure and approximating the portfolio dynamics with a scenario tree, we formulate the portfolio optimization problem into a multi-stage stochastic linear program. Each E&P project covers the full life cycle of an unproven prospect and is modeled by a decision tree with mixed-integer techniques. The optimization model coordinates multiple trees by a shared account and inventory-type budget constraints. All incurred cash flows from the projects are either deposited into or drawn from the account. Any excess of the account is carried to the next period like an inventory, earning interests at the risk-free rate. We provide extensive numerical experiments to show the effectiveness of the model. We also discuss possible extension to make the model more practical.

In Chapter 5, we develop a scenario generator to generate the scenario tree for the multistage portfolio optimization model developed in Chapter 4. Since E&P projects are long-term investments, the investor has chances and options to buy information to make educated guesses about the future portfolio performance. Sensible learning comes from reliable statistical dependences. In this chapter, we identify two types of project dependences, inter-project dependence and intra-project dependence, to support proper statistical learning. We use sequential statistical learning to progressively update the investor's belief about the portfolio performance. The scenarios are then generated from the updated distributions. To obtain a solvable multistage stochastic program, we incorporate different scenario reduction schemes in the scenario generation algorithm to restrict the growth of the multistage scenario tree.

Finally, Chapter 6 summarizes the key results and major contributions and limitations, and gives an outlook on future research.

Chapter 2 Oilfield exploration and production

This dissertation is about creating optimization models to manage the risk of exploration and production (E&P) projects. Before we delve into the modeling details, we would like to spend this chapter on basic knowledge about the E&P business and methods to assess E&P project value and manage project risk. This chapter would serve as a short background knowledge summary to those who do not have experience in the upstream business and/or corporate finance.

This chapter is arranged as follows. In §2.1, we give a brief review to the E&P activities and present a phase-model to cover the life cycle of typical E&P projects. In §2.2, we collect and present some statistics about E&P projects, which would give us a basic idea about the economic side of E&P projects. In §2.3, we demonstrate how to use tank model to forecast well or reservoir production performance. In §2.4, we introduce the fundamental ideas to assess E&P projects from corporate finance and investment science courses.

2.1 A Brief Review about Exploration and Production

In this section we give a brief review on E&P projects to make this dissertation a self-contained one. We cover the major issues from how oil was formed to how oil is found and produced. The review is based on Lyons (1996), Walsh and Lake (2003), and the official website of Energy Information Administration (EIA) of the U.S. Department of Energy (http://www.eia.doe.gov).

2.1.1 Oil rocks, traps, and prospects: where oil comes from

Crude oil and natural gas together are called petroleum. Petroleum is a mixture of hundreds of different hydrocarbon compounds whose molecules contain only hydrogen and carbon. Petroleum and coal are the most widely used fossil fuels.

Petroleum was formed from ancient animal, plant, and marine life remains. They were washed into lakes or seas and then buried by layers upon layers of sediments. Over millions of years, thick sediment layers applied intense heat and pressure to convert the organic remains into petroleum in the form of liquid (crude oil) or vapor (natural gas).

Oil and gas are not stored in big and black underground pools. Instead, they exist as tiny droplets trapped in underground rocks, called oil rocks. For rocks to contain oil and gas, they must have two properties. First, they must contain open space, called "pores," to store oil and gas. The number of pores in a rock determines how much oil and gas can be stored. Second, those pores must be interconnected and large enough to allow oil and gas to flow through the rock. Those two properties are measured by porosity and permeability, respectively. Porosity measures the density of pores in a rock and permeability quantifies the size and inter-connectivity of pores. Shale, sandstone, and limestone are well known porous sedimentary oil rocks.

Geologists need examine core samples to estimate the presence of commercial oil. Besides rock properties like porosity and permeability, geologists also need inspect properties of trapped oil droplets such as viscosity which measures the resistance of oil to flow through pores or cracks.

Underground petroleum is under enormous pressure imposed by millions of tons of rocks lying on it and evaporated gases boiled by natural heat. Like any liquid under pressure, oil droplets squeeze into pores or tiny cracks of rocks and migrate toward the earth surface (the direction to release pressure) until they are trapped in a porous and permeable layer of rocks, called a trap, by an impermeable barrier, called a seal, over the trap. A trap is called a prospect, or an oil reservoir, if the trap contains sufficient petroleum to commercialize. If the seal is missing, oil might have risen through rocks to the surface and left nothing behind over millions of years. Besides the function of the seal to prevent oil leaking out of the reservoir rocks, the seal also preserve the accumulated pressure which contributes to the commerciality of the reservoir.

The typical shape of the seal is like an upside-down bowl covering the reservoir trap. However, due to tectonic movements, the original seals and reservoirs may have been broken into parts, transported and shifted away, and folded into various shapes, which makes exploratory activities to verify and locate hydrocarbons highly unpredictable.

In summary, to guide the exploration, people have developed a set of theories to explain how petroleum is formed, where it migrates to and from, where to find it, and what the trap mechanism is. Once we have sufficient information to answer those questions, we should have better idea about the presence and commerciality of underground oil.

2.1.2 Acquisition and exploration

In the distant past, finding an oil prospect was mainly a matter of luck. The only reliable clue is observable oil seeping to the surface. After more than one century of development, the current technologies and data analysis techniques have significantly improved the likelihood to find commercial oil.

The E&P activities start with acquiring the right to explore a field. In the United States, the Bureau of Ocean Energy (BOE), formerly known as the Minerals Management Service (MMS), of the Department of Interior annually schedules lease auctions and sets rules for companies to bid new offshore areas. Prior to the auction, the BOE takes up to five years to publish pre-sale lease drafts, collect public comments, and prepare final leasing plans. In the mean time, oil companies review the announced areas and assess whether to compete with other players to bid for any potential blocks while taking into account their current holdings, reserve replacement goals, and production growth plans.

If a company is interested in some block, it will assign a team of geologists to collect public information about the block and probably also purchase private maps and data. In some cases, the company may be allowed to take exploratory tests on the existence of hydrocarbons in preparation for the bid. Based on the initial information, the experienced geologists develop a complete story for the block about the oil origin, migration history, reservoir rock formation, and trap mechanism. They use the story to convince the management of the company to commit real money to acquire the right to explore the field. If the story is solid and consistent to the company's strategic goals, the company will submit a sealed bid to the BOE. The bid components include bid price, royalties, and other commitments. The lease term for initial explorations is normally between five to ten years, depending on the geological conditions such as depth.

After a successful bid, the company will carry out a series of exploration activities including seismic surveys, wildcat drilling, and core sample studies to collect more accurate information to identify the locations of potential prospects in the block.

Seismic survey studies the collected shock waves reflected from underground rock layers. When man-made shock waves are imposed to the earth surface, the shock waves travel downward at different speeds through different types of rocks. Part energy of the shock waves are reflected while crossing boundaries between different rocks. Experienced geologists can identify subsurface formations and their locations by reading shock waves reflected at different angles and intervals. With the help of computer scientists and mathematics, the geologists can visualize the subsurface structures as interactive three dimensional virtual reality images.

Other modern exploratory techniques include telemetric imaging such as aerial or satellite images, detecting geomagnetism or gravity fields variations caused by underground oil volumes, and so on. These modern exploratory techniques discover important information about whether proper conditions and formations exist in underground rock layers which may trap commercial oil or gas. The technologies have not only significantly increased the chance of success to drilling wildcat wells, but also reduced wastes and improved capital efficiency.

For deepwater explorations, one or more exploratory wells, called “wildcat wells,” must be drilled to verify the findings based on seismic surveys or other preliminary explorations. Rock core samples will be taken from preset depths to check properties of reservoir rock and to see whether the targeted rock layers are likely to contain oil. Since drilling an exploration well can be very expensive, up to several hundred million dollars, the drilling decisions must be made with enough accurate information to make a rewarding exploration.

2.1.3 Delineation and facility construction

Once the existence of a potential prospect is identified by seismic surveys and exploratory drilling, the size and boundary of the prospect must be further delineated by drilling appraisal wells. The geologists will assess the appraisal results to see whether it is worthwhile to take risks to commercialize the reservoir. If the result is unfavorable due to, for example, insufficient reserves or low pressure, the prospect will be deferred or abandoned. As a result, the company loses all the expenditures. Otherwise, if the appraisal exploration reveals positive signals, the geologists will propose the company to move on to the production preparation stage.

In the preparation stage, a development team will be recruited to carry out concept studies. A concept is characterized by the capacity of the proposed platform and the flexibility to increase the capacity in subsequent production expansions. The team is responsible to determine the proper concept for the future depletion of the prospect.

Sometimes, the production wells of the new prospect can be drilled from nearby platforms. Therefore, new wells can share the same platforms and processing capacities with existing wells. However, if the new prospect is not close to any platform or the expected production rates of the new prospect significantly exceed the residual processing capacities of nearby platforms, additional platforms and associated facilities must be constructed prior to production. In this case, the decision where to locate the new platform becomes a strategic one since the new platform may cover newly acquired areas in the future. The location-related problems have been extensively studied in OR society and earned names as the (capacitated) facility location problem and/or the set covering problem. In our research, we will not consider the location decisions. Instead, we are more interested in the selection of concept, i.e, the selection of offshore platform types and capacities.

Depending on the size, depth, and other configurations of targeted reservoirs, the platforms can be divided into three categories, fixed platforms, floating systems, and subsea systems. The following Figure 1-1 shows examples for some typical platforms. Fixed platforms are concrete or steel structures physically grounded onto the sea floor, supporting a deck with space for drilling rigs, production equipment, and crew quarters. The leftmost three platforms in Figure 1-1 are three types of fixed platforms. All the remaining platforms, excluding platform numbered 10, are floating platforms. Floating platforms are, by their names, structures floating or semi-submersible on the surface of lake or sea, softly moored or tethered in place with rope and chain. The last platform in Figure 2-1 connects to a series subsea wellheads with pipes. Subsea systems refer to wellheads sitting on the sea floor and connecting directly to a host (fixed or floating) platform or to a subsea manifold by pipelines. Subsea systems may not only save capital expenditures by avoiding the construction of additional platforms, but also help people to decrease environmental impact. The environmental concern has increasingly become a substantial factor determining whether a prospect can be brought to final production.

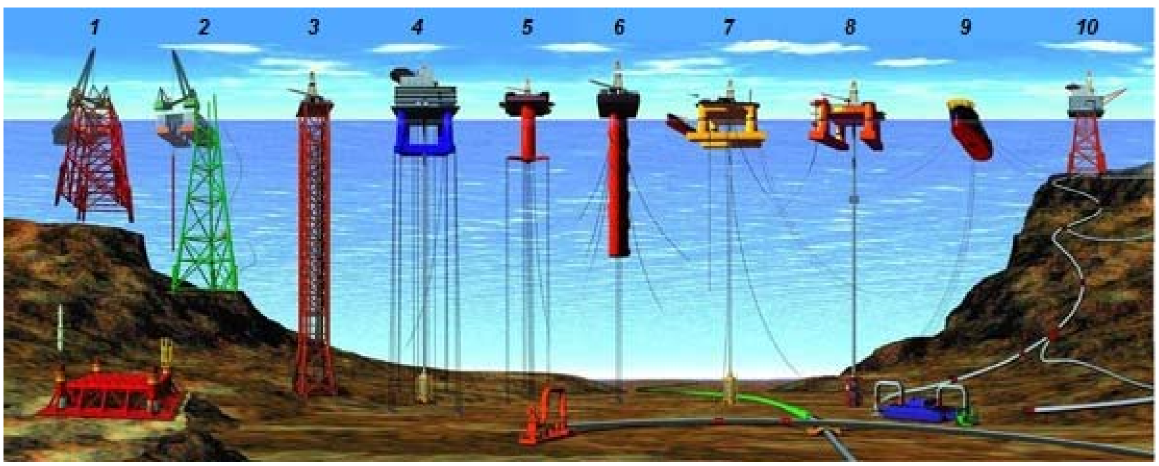


Figure 2-1: Types of offshore oil and gas structures include: 1, 2) conventional fixed platforms; 3) compliant tower; 4, 5) vertically moored tension leg and mini-tension leg platform; 6) Spar; 7,8) Semi-submersibles; 9) Floating production, storage, and offloading facility; 10) sub-sea completion and tie-back to host facility. (Courtesy: NOAA[1].)

Leffler et al. (2003) provide an excellent description to major oil platforms and their application scenarios. Additional information about oil platforms can be found on

[1] http://oceanexplorer.noaa.gov/explorations/06mexico/background/oil/media/types_600.jpg

the official website of National Oceanic and Atmospheric Administration (NOAA) and the free online encyclopedia, Wikipedia.

During the process of conceptual studies, the development team and the reservoir management team will work together to draft the construction plans and probably a rough drilling schedule. With the draft, the company needs to compile and submit a development and production proposals to the Interior to acquire necessary permits to bring the reservoir into production. The company can not take further development actions until the proposals are approved and the permits are authorized. Once the company is permitted to move to the next phase, the company will carry out the construction plan to build production platforms and facilities.

2.1.4 Production (primary, secondary, and tertiary recoveries)

In the course of facility construction and pipeline connection, a team of drilling engineers will prepare the drilling prognosis to detail the drilling schedules. Meanwhile, the development team will prepare the detailed production plans. Ideally, the company should make plans to get oil out of the ground as soon as possible to shorten the payback period and maintain sufficient cash reserves. However, more often than not, the actual drilling and production activities may be deferred or changed in accordance to the company's varying short term and long term goals and the volatile market conditions. In the uncertain environment, the flexibility to accordingly change plans gives the company additional values. Smith and Mccardle (1999) show a real E&P project whose value is substantially affected by oil prices and the inclusion of options to defer drilling.

The life of a typical production well can be divided into five phases, planning, drilling, completion, production, and abandonment. The drilling team prepares the well plan which includes the shape, orientation, depth, completion, and evaluation. The team also propose the drilling prognosis which splits the targeted production well depth into a series of consecutive depths. When the drill bit reaches a designated depth, a section of increasing narrower steel pipe (casing) will be installed with cement being filled between the outside of the casing and the borehole. The casing and cement together provide structure integrity to the wellbore and also isolate cross-sectional zones of different high pressures. During drilling, drilling mud will be pumped down through the casing to circulate bottom rock cuttings back to the surface. When the drill bit reaches the targeted depth, well completion is performed to create fractures in the production zone so that oil droplets can flow into the production tubing. If the natural pressure at the bottom is high enough, a production tree, also called Christmas tree (a set of integrated valves to regulate flow pressure, also called), is sufficient to maintain the production. Otherwise, artificial lift methods will be used. Sometimes, the extracted water and natural gas will be separated from oil and injected back to maintain reservoir pressure through nearby retired wells. The final step before production is to connect the outlet valve of the well to storage tanks which are connected to a distribution network with pipelines.

For a new oil field with no or short production history, the natural pressure in the reservoir is normally high enough for the oil or gas to flow to the surface. As long as the pressure in the reservoir remains high enough, the production can continue with natural pressure as the only drive. The production phase without artificial lifts is called primary recovery. Since water coexists with petroleum in porous oil rocks, the output from the wells is a mixture of crude oil, natural gas, and water and must be separated.

All underground physical substances are compressible under the great pressure of millions of tons of rocks above them. With the resources flowing out up the wells, the volume of any underground substance will expand as the pressure is reduced.

As the production continues, the reservoir pressure starts dropping at certain time when the fluids are depleted. When the pressure drops down to the level that continuing production is not economical, the operator needs to make a decision whether to abandon the wells or to apply certain artificial lift methods to increase the reservoir pressure and the production rate. Common artificial lifts include water-flooding, steam flooding, and gas injection. The purposes are to use artificial efforts to respectively wash out oil droplets from rock pores, introduce heat to increase oil droplets liquidity, and inject gases which expand and increase the reservoir pressure with the aid of underground heat. The production with the aid of artificial lift methods is called secondary recovery.

After the primary and secondary recovery, there is a third recovery called tertiary recovery which involves injecting detergent-like chemical compounds to improve the effectiveness of water-flooding.

According to the Department of Energy of U.S., the primary recovery rate is about 25%, the secondary recovery rate is between 5% and 10%, and the tertiary recovery rate is far less than 1%. So the primary recovery and the secondary recovery together produce about 30~35% of total original oil in place (OOIP), and almost two-thirds (2/3) oil has been or will be left behind under current technologies. In other words, whenever we produce one barrel oil we leave two barrels intact underground. The fact also tells us that when people mention a well is dry or a reservoir is depleted, they don’t really mean there is no oil or gas. They are actually saying that under current technique conditions it is not economic to bring the remaining underground oil to the surface of the earth.

2.1.5 Abandonment

Secondary recovery may extend the life of a well or a reservoir. However, as the reservoir is depleted, the production rate will finally drop to some low level that it is no longer profitable to continue the production. The level is called economic limit. A well is said to reach an "economic limit" when its production rate doesn't cover the overall expenses, including royalties and taxes. It can be determined by the following factors: royalties (ρ, %), taxes (α, %), lease operating cost (OPEX, \$/barrel), and current oil price (P, \$/barrel). If we set the after tax income >= to the lease cost we obtain:

$$p \cdot q(1-\rho)(1-tax) - WI \cdot LOE \geq 0 \quad (2.1.1)$$

Solving for the q value which satisfies the above as an equality:

$$q_{eco} = \frac{WI \cdot LOE}{p \cdot (1-\rho)(1-tax)} \quad (2.1.2)$$

2.2 Reservoir Performance Prediction: The Tank Models

To assess the performance of E&P projects under different operations, we need a model to describe their production flows as a function of geological parameters and operational decisions. The model should be simple and flexible so that we can conveniently incorporate it in optimization models to find optimal development plans without causing invincible computational challenges. For simplicity, the model we are looking for will be restricted to only cover the primary recovery.

There are five popular methods to predict reservoir productions. In order of increasing sophistication, they are (1) intuition and judgment, (2) analog, (3) empirical correlations, (4) decline-curve, and (5) numerical simulation. The characterizations and comparisons of these methods are summarized in Walsh and Lake (2003). The last one is our choice due to its flexibility and consistence to describe the multi-stage field development.

The most popular numerical simulation models are tank models. Tank models are a large class of numerical models tailored reservoirs of different types. The tank model we will employ is for a special reservoir type which contains a single-layer of homogenous compressible liquid. In the tank model, the production is driven by pressure release due to compressible hydrocarbons' expansion toward wellbores, through well pipes, and up to the surface. The model mainly captures the primary recovery. The tank model can be applied to a single well, a group of wells, or a whole reservoir. The main results of the tank model are, if the production is only driven by the natural reservoir pressure, i.e. the primary recovery, the pressure declines exponentially and, accordingly, the reservoir production rate declines exponentially as well.

In this section, we first give a brief review to the classic tank model of single-layer compressible liquids based on Walsh and Lake (2003). Next, we extend it to a dynamic capacitated tank model which allows sequential drilling or shut-in decisions and conforms capacity restrictions. The dynamic model is generalized from the numerical example in Hultzsch (2005). Finally, since the tank model itself conveys high degree of nonlinearity in operational decisions, we propose a linearization scheme for the dynamic tank model using mixed integer programming (MIP) techniques.

The tank model directly deals with geological parameters and operational decisions. So before we step into the model mathematics, we give the definitions and units of relevant reservoir properties and operational parameters. The reservoir properties are categorized into three classes, the properties for reservoir rock, for oil, and for reservoir including rock and oil as a whole body.

Table 2-1: Fundamental geological properties for the tank model.

	Name	Definition	Unit[2]
Reservoir rock	Permeability w.r.t. oil (k_o)	It is the ability of the rock to conduct oil through its interconnected pores.	mD
	Porosity (ϕ)	It is the percentage of the rock volume that can contain fluids.	-
Hydrocarbons	Oil viscosity (μ_o)	It measures the resistance of oil to flow.	cp
	Oil saturation (S_o)	It is the oil percentage of the fluids (oil, gas, water) trapped in the rock pores.	-
	Compressibility factors (c_t, c_o, c_w, c_f)	They measure the compressibility of the reservoir rock and the extracted fluids (oil and water). The overall compressibility of the reservoir can be calculated as: $c_t = S_o c_o + (1 - S_o) c_w + c_f$, where c_o, c_w, and c_f are the compressibility of oil, water, and formation rock respectively.	psi^{-1}
	Oil formation volume factor (B_o)	It accounts for the difference between the volume of the same amount of oil as measured at the surface conditions and at reservoir conditions.	rb/STB
Reservoir /	Drainage area (A)	It is the area of the reservoir drained by the well(s).	acre
	Reservoir thickness (h)	It measures the vertical height of the net pay. In geology, pay refers to a reservoir or portion of a	ft

[2] In oil industry, the convention is to use unit prefixes *m*, *M*, and *MM* to stand for the magnitudes 10^{-3}, 10^{3}, and 10^{6}, respectively.

		reservoir that contains economically producible hydrocarbons.	
	Skin factor (s)	It models the difference from the pressure drop predicted by Darcy's law due to the skin (a thin region) around the wellbore that causes permeability damage.	-
	Reservoir pressure ($\overline{p}$)	It is the pressure of fluids within the pores of a reservoir exerted by the rocks and water supported by the reservoir. The reservoir pressure measured in a discovery well is said the initial pressure of the reservoir, denoted as $\overline{p}_i$.	psi

Table 2-2: Operational parameters for the tank model.

Name	Definition	Unit
Well radius (r_w)	It indicates the radius of the wellbore.	Ft
Shape factor (C_A)	It takes care of the geometry of the drainage area for each producer (single well). When it takes a circular shape, Walsh and Lake (2003) give $C_A = 31.62$.	-
Bottom-hole pressure (p_{wf})	It is the pressure measured at the bottom of a production well.	psi
Number of wells (N_w)	It is the number of production wells.	-
Economic limit (q_{el})	It is the lowest commercial production rate.	bbls/day

2.2.1 Basic tank model

The tank model describes the reservoir dynamics by setting up connections among the production rate, pressure, and the rate of pressure drop. The connections are based on the law of conservation of mass and Darcy's law in slightly-compressible liquid.

Let's consider a single production period $[t_0,\ t_1]$ with $t_0 = 0$ and $t_1 = T$, the life length of the project. We assume that there is no production history prior to time t_0. At time t_0, N_w production wells are drilled and start production. The reservoir pressure at t_0 equals to the initial reservoir pressure, $\overline{p}(t_0) = \overline{p}_0$, which is uncovered by wildcat drilling. We assume no major operational changes occur since then, i.e., no shut-ins and no new drillings. We use $\overline{p}(t)$ and $q(t)$ to denote the average reservoir pressure and reservoir production rate at time $t \in [t_0, t_1]$, respectively. If we denote the production rate of well *k* of the reservoir at time *t* as $q_k(t)$, the total reservoir production rate at time *t* is $q(t) = \sum_k q_k(t)$.

The tank model says the production rate of well *k* at any time *t* is proportional to the pressure difference between reservoir and the bottom hole at well *k,*

$$q_k(t) = \frac{J_k\left[\overline{p}(t) - p_{wf,k}\right]}{B_0} \quad \text{(STB/day)}, \qquad (2.2.1)$$

where J_k is the (daily) productivity index (PI) of well *k*, $\overline{p}(t)$ and $p_{wf,k}$ the reservoir pressure and the bottom hole pressure of well *k*, and B_0 the formation volume factor which converts the unit of production rates from Reservoir Barrel (rb) to Standard Tank Barrel (STB). The index J_k measures the performance of well *k* as of how much oil (in rb) it can produce per day using per unit of the pressure difference. It is estimated by

$$J_k = \frac{0.00708\, h\, k_0}{\mu_0 \left(\frac{1}{2} \ln \frac{A_k}{r_{w,k}^2 C_A} + 5.75 + s \right)} \quad \text{(rb/psi/day),} \tag{2.2.2}$$

where A_k is the well k's drainage area and the meaning of other parameters please refer to Table 2-1 and Table 2-2. In theory, each of those parameters in the above two formulae is allowed to vary from well to well.

The production continues as long as the reservoir pressure is high enough to maintain production at a rate exceeding the economic limit q_{el}. We assume the bottom-hole pressure $p_{wf,k}$ is constant during the production. The minimal reservoir pressure $\overline{p}_{\min}$ required to sustain the commercial production is given by $\frac{J_k \left[\overline{p}(t) - p_{wf,k} \right]}{B_0} \geq q_{el}$, which leads to

$$\overline{p}_{\min} = \frac{B_0 \cdot q_{el}}{J_k} + p_{wf,k} \quad \text{(psi).} \tag{2.2.3}$$

If the reservoir pressure is greater than $\overline{p}_{\min}$, the well can be economically operated.

Both the rock and the fluids in the reservoir are compressible and under enormous pressure. During production, they expand to release the reservoir pressure, driving the fluids to flow to the surface through production wells. The law of conservation of mass leads to the material balance conditions as the following formula (2.2.4),

$$-\frac{V_p c_t}{B_0} \frac{d\overline{p}(t)}{dt} = q(t). \tag{2.2.4}$$

The formula says that during the production, the rate of the reservoir pressure is proportional to the reservoir depletion (or production) rate. The details how the formula comes from please refer to Walsh and Lake (2003, p.294). In (2.2.4), V_p is the pore volume of the reservoir and c_t is the total compressibility of the rock and fluids contained in the reservoir. The pore volume indicates the available space to contain fluids. It can be calculated by

$$V_p = 7757.792 A h \phi \quad \text{(rb)}, \tag{2.2.5}$$

where A is reservoir drainage area (in acres), h is reservoir thickness (in feet), and ϕ is the porosity (in %) of the reservoir. The fluids trapped in the pore space may be a mixture of oil (hydrocarbons) and water. We use the oil saturation factor S_o to specify the percentage of oil mixed in the reservoir fluids. The water saturation factor is defined accordingly by $S_w = 1 - S_o$. So the original oil in place (OOIP) is given by

$$V_{OOIP} = \frac{V_p\, S_o}{B_o} = \frac{V_p\left(1 - S_w\right)}{B_o} \quad \text{(STB)}. \tag{2.2.6}$$

From (2.2.1), (2.2.4), and $q(t) = \sum_k q_k(t)$, we can get the exponential decline model for both the reservoir production and pressure by integrations in proper time and pressure spaces, respectively,

$$\lambda = \frac{\sum_k J_k}{V_p\, c_t} \quad \text{(1/day)}, \tag{2.2.7}$$

$$\bar{p}(t) = \bar{p}_{wf} + \left(\bar{p}_0 - \bar{p}_{wf}\right) e^{-\lambda(t - t_0)} \quad \text{(psi)}, \tag{2.2.8}$$

$$q_k(t) = q_k(t_0) e^{-\lambda(t - t_0)} \quad \text{(STB/day)}, \tag{2.2.9}$$

$$q(t) = \sum_k q_k(t) = q_0 e^{-\lambda(t-t_0)} \quad \text{(STB/day)}, \tag{2.2.10}$$

where $\overline{p}_{wf} = \dfrac{\sum_k J_k \cdot p_{wf,k}}{\sum_k J_k}$ (psi), $q(t) = \sum_k q_k(t)$ (STB/day), and $q_0 = \sum_k q_k(t_0) = \dfrac{\sum_k J_k \left[\overline{p}_0 - p_{wf,k}\right]}{B_0}$ (STB/day). So far, all parameters are measured in days and the time index t is in days. We can annualize these parameters so that $\lambda = \dfrac{365 \sum_k J_k}{V_p\, c_t}$ (1/year), $q_0 = \dfrac{365 \cdot \sum_k J_k \left[\overline{p}_0 - p_{wf,k}\right]}{B_0}$ (STB/year), and the time index t is measured in years.

For simplicity, we ignore location variations and assume all wells are perfectly communicable to the reservoir and have identical parameters. In this case, we can calculate these parameters by picking an arbitrary well k as a representative and apply its parameters to all other wells. As a consequence, we have $A_k = A/N_w$, $\sum_k J_k = N_w J_k$, $\overline{p}_{wf} = p_{wf,k}$. We can rewrite the above formulas in the following annualized form,

$$q_0 = N_w q_k(t_0) = \frac{365 N_w J_k \left[\overline{p}_0 - p_{wf,k}\right]}{B_0} \quad \text{(STB/year)}, \tag{2.2.11}$$

$$\lambda = \frac{365\, N_w J_k}{V_p\, c_t} \quad \text{(1/year)}, \tag{2.2.12}$$

$$q(t) = q_0\, e^{-\lambda(t-t_0)} \quad \text{(STB/year)}, \tag{2.2.13}$$

$$\overline{p}(t) - \overline{p}_{wf} = \left(\overline{p}_0 - \overline{p}_{wf}\right) e^{-\lambda(t-t_0)} \quad \text{(psi)}, \tag{2.2.14}$$

2.2.2 Capacitated tank model

The classic tank model is un-capacitated and static. In this section, we consider the facility capacity due to storage space, pipeline size, processing capability, and so on.

In the next section, we present a dynamic tank model which not only incorporates the capacity constraints, but also allows multistage drillings.

The reservoir pressure decline is caused by depletion. Since hydrocarbons flow continuously out of the reservoir, the reservoir pressure should change continuously as well. We shall see later that in the dynamic tank model, the production rate as a function in time is smooth and integrable almost everywhere except for the epochs where drilling decisions are made. Let $Q(t)$ denote the accumulated production starting at t_0 and up to time $t \geq t_0$. By definitions, we have $Q(t) = \int_{t_0}^{t} q(u)\,du$. Integrating both sides of (2.2.4) over a short time period $[s,t] \subseteq [t_0, T]$ gives us the formula as below to predict the reservoir pressure based on the accumulated production during the period,

$$\bar{p}(t) = \bar{p}(s) - \frac{B_0}{V_p c_t} \cdot \left[Q(t) - Q(s) \right] \text{ (psi).} \qquad (2.2.15)$$

Next we extend the tank model to predict reservoir production with the consideration of the processing capacity. To be consistent with the later developed model, we consider the dynamics of reservoir production and pressure during a production period $[s,t]$ of one year long, i.e., $\Delta t = t - s = 1$ (year). We denote the annual processing capacity for the reservoir as C (STB/year). When no operational changes occur in the period $[s,t]$, the reservoir produces either at its annual processing capacity or at natural flow which is described as the exponential decline model based on (2.2.13), $q(u) = q(s) \cdot \exp\left[-\lambda_s \cdot (u - s) \right]$, $u \in [s,t]$, where the decay rate λ_s is determined by the number of active producing wells N_s at time s. Consequently, we can compute the cumulative production during the year as

$$Q(t)-Q(s)=\min\left\{\int_s^t q(u)\,du,\ C\cdot\Delta t\right\}$$
$$=\min\left\{\frac{q(s)-q(t)}{\lambda_s}=q(s)\frac{1-e^{-\lambda_s\cdot\Delta t}}{\lambda_s},\ C\cdot\Delta t\right\} \text{ (STB).} \qquad (2.2.16)$$

The first term of (2.2.16) corresponds to the declining natural production when the capacity is not reached. The second term corresponds to the constant production at the capacity when the annual production would exceed the capacity C if the capacity restriction is removed. Combining (2.2.15) and (2.2.16) gives us the way to compute reservoir pressure.

2.2.3 Dynamic tank model

The static model is insufficient to capture the reservoir's dynamic behavior when wells are drilled sequentially. In this section, we develop a dynamic tank model with the considerations of capacity constraints and expansion. The model is a generalization of the numerical example of Hultzsch (2005).

We consider a dynamic tank model in which annual drilling or shut-in decisions are made at the beginning of each year. For simplicity, new wells are assumed to start production immediately after the drilling decisions. We consider discrete time points t = 0, 1, 2, …, T. Each time point t is a decision stage at which either additional new wells are drilled or some production wells are shut in. We assume the relationship (2.2.4) is still valid during each period, although some parameters of the tank model may change at the beginning of the period due to the operational changes. The dynamic tank model relies on one critical assumption that the reservoir pressure continuously drops as long as the fluids depletion continues. So the reservoir pressures are the same immediately before and after any operational change, although the production rate and the pressure drop rate may jump.

We generalize the continuous time static tank model's formulae (2.2.11) ~ (2.2.14) to discrete time points $t = 1, 2, \ldots$. For each stage t, we use the positive number n_t to denote the number of producing wells during the immediate following period. If $n_t > n_{t-1}$, then $n_t - n_{t-1}$ new wells are drilled at stage t. Similarly, $n_t < n_{t-1}$ implies $n_{t-1} - n_t$ producing wells are shut in at t. Let $y_t = \int_t^{t+1} q(u)\,du$ be the production during the period $[t,\ t+1)$, i.e., $y_t = Q(t+1) - Q(t)$. We extend the notations of the static tank model to the dynamic situation by adding stage subscript t to relevant parameters as follows, given initial pressure $\bar{p}_0$

$$J_t = \frac{0.00708\,h\,k_0}{\mu_0\left(\frac{1}{2}\ln\frac{A}{r_w^2 C_A n_t} + 5.75 + s\right)} \quad \text{(rb/psi/day)}, \tag{2.2.17}$$

$$\lambda_t = \frac{365\,n_t J_t}{V_p\,c_t} \quad \text{(1/year)}, \tag{2.2.18}$$

$$q_t = n_t q_k(t) = n_t \frac{365 \cdot J_t\left[\bar{p}_t - \bar{p}_{wf}\right]}{B_0} \quad \text{(STB/year)}, \tag{2.2.19}$$

$$y_t = \min\left\{ q_t \frac{1 - e^{-\lambda_t \cdot \Delta t}}{\lambda_t},\ \ C \cdot \Delta t \right\} \quad \text{(STB/year)}, \tag{2.2.20}$$

$$\bar{p}_{t+1} = \bar{p}_t - \frac{B_0}{V_p c_t} \cdot y_t \quad \text{(psi)}, \tag{2.2.21}$$

where the subscript of c_t in (2.2.18) and (2.2.21) is an exception and it is an abbreviation for "total."

Given n_t producing wells during the period $[t,\ t+1)$, formulae (2.2.17) ~ (2.2.19) calculate the values of three corresponding parameters determined at the beginning of the period. The formula (2.2.20) comes from (2.2.16) and estimates the annual production during the period. The annual production should not exceed the annual processing capacity *C*. The formula (2.2.21) estimates the reservoir pressure drop during the production period and calculates the pressure at the beginning of the succeeding period. Iteratively applying the formulae (2.2.17) ~ (2.2.21) can fully describe the reservoir dynamics as a function of multistage drilling/shut-in plans $\{n_t\}_{t=0}^{T-1}$.

Unfortunately, the above dynamic tank model is highly nonlinear in the drilling decisions. In order to assess oil projects of realistic sizes with optimization methods, it would be of great help if we can linearize the tank model. This is the topic of the next section.

2.2.4 Linearization of the dynamic tank model

The first step of our linearization is to get a more compact formulation of the dynamic tank model by plugging (2.2.19) in to the first term of the right hand side of (2.2.20). Noticing (2.2.18), we can combine (2.2.19) and (2.2.20) into just one formula (2.2.24) as follows. With other formulae being rewritten again, the resulting complete dynamic tank model is as follows, for *i* = 0, 1, 2, …, *T*-1.

$$J_t = \frac{0.00708\, h\, k_0}{\mu_0 \left(\frac{1}{2} \ln \frac{A}{r_w^2 C_A n_t} + 5.75 + s \right)} \quad \text{(rb/psi/day)}, \tag{2.2.22}$$

$$\lambda_t = \frac{365\, n_t J_t}{V_p\, c_t} \quad \text{(1/year)}, \tag{2.2.23}$$

$$y_t = \min\left\{ \frac{V_p\, c_t}{B_0}\left(1 - e^{-\lambda_t \cdot \Delta t}\right) \cdot \left(\overline{p}_t - \overline{p}_{wf}\right),\ C \cdot \Delta t \right\} \quad \text{(STB)}. \tag{2.2.24}$$

$$\overline{p}_{t+1} = \overline{p}_t - \frac{B_0}{V_p c_t} \cdot y_t \quad \text{(psi)}. \tag{2.2.25}$$

From formulae (2.2.22) and (2.2.23), the productivity indices and the decay rates are solely determined by the number of producing wells and do not rely on the production history. So we can calculate two long lists of values $\{J_n\}_{n=1}^{n=N_{\max}}$ and $\{\lambda_n\}_{n=1}^{n=N_{\max}}$ for both parameters as a function in the number of wells before hand, where $N_{\max}$ is the maximal number of wells that can be drilled in the reservoir. Therefore, we can calculate a long list of the coefficient $\frac{V_p\ c_t}{B_0}\left(1-e^{-\lambda_t \cdot \Delta t}\right)$ in the first term of the right hand side of (2.2.24) as a function in *n*, the number of wells. Consequently, we can model the drilling or shut-in decisions by the 0-1 multiple choice constraints and SOS1 type variables of mixed integer programming techniques.

Next, we formulate the multistage deterministic oilfield development problem as a mixed integer program.

Problem statement

The model is tailor from Hultzsch (2005)'s numerical example by adding capacity constraints and other restrictions. Assume we have a proven reservoir whose geological parameters are known, we have T years to develop it. We can drill new wells or shut in producing wells at the beginning of each year, but not during the interims. At any time, the active wells can not exceed the maximal allowed number of wells $N_{\max}$ of the reservoir, due to regulations to protect the environment. The project life of T is divided into two phases. In the first phase, the reservoir produces at the initial production capacity C. In the second phase, the capacity is expanded by installing additional capacity C_exp. Finally, at any time, if the reservoir pressure drops to the minimal economic production level $\overline{p}_{\min}$, the reservoir is not commercially viable any more and will be abandoned.

Given above conditions and restrictions, our objective is to find the optimal development plans to maximize the total cumulative production. The production of the reservoir is predicted by the capacitated dynamic tank model (2.2.22) ~ (2.2.25).

Indices and index sets

$\mathbb{T}$: the set of decision time points, $\mathbb{T} = \{0, \cdots, T-1\}$, indexed with *t*.

$\mathbb{T}_1$, $\mathbb{T}_2$: the sets of decision points for the two production phases correspond to initial capacity and expanded capacity, respectively, and $\mathbb{T} = \mathbb{T}_1 \cup \mathbb{T}_2$.

$\mathbb{K}$: the set of operating plans, $\mathbb{K} = \{0, \cdots, K\}$, indexed with *k*. The index of each plan indicates the number of production wells during a production period.

Decision variables

$$x_{tk} = \begin{cases} 1, & \text{the number of active wells during } [t, t+1) \text{ is } k \\ 0, & \text{the number of active wells during } [t, t+1) \text{ is not } k \end{cases}.$$

y_t (STB): the total production during the period $[t, t+1)$.

y_ub_t (STB): the maximal production potential during the period $[t, t+1)$.

p_t (psi): the reservoir pressure at the beginning of period $[t, t+1)$ or at the end of period $[t-1, t)$ when the period is valid.

Parameters

C (STB/year): the initial annual processing capacity for the reservoir.

C_exp (STB/year): the expanded annual processing capacity for the reservoir.

$B = \dfrac{B_0}{V_p c_t}$ (psi/STB): the coefficient preceding y_t in (2.2.25).

$\overline{p}_{\min}$ (psi): the minimal pressure required to maintain production, as in (2.2.3).

M: the big-M factor, usually to be some large enough number.

$A_0 = 0$ (STB/psi): the coefficient preceding $\left(\overline{p}_t - \overline{p}_{wf}\right)$ in (2.2.24) corresponds to the abandonment case where no well is active during $\left[t, t+1\right)$, or $k = 0$.

$A_n = \dfrac{V_p\ c_t}{B_0} \cdot \left(1 - e^{-\lambda_n}\right)$ (STB/psi): the coefficient preceding $\left(\overline{p}_t - \overline{p}_{wf}\right)$ in (2.2.24) corresponds to n active wells during $\left[t, t+1\right)$ for $k \in \mathbb{K}$. It depends on the following two implicit parameters,

$$J_k = \frac{0.00708\, h\, k_0}{\mu_0 \left(\dfrac{1}{2} \ln \dfrac{A}{r_w^2 C_A \cdot k} + 5.75 + s \right)} \quad \text{and} \quad \lambda_k = \frac{365 \cdot k \cdot J_k}{V_p\ c_t}, \text{ for } k \in \mathbb{K}.$$

The optimization model

$$z^* = \underset{\{x, y, y\_ub, p\}}{\text{Max}} \quad \sum_{t \in \mathbb{T}} y_t \qquad (2.2.26)$$

$$\text{s.t.} \quad y_t \le C, \qquad \forall t \in \mathbb{T}_1 \qquad (2.2.27)$$

$$y_t \le C + C\_exp, \qquad \forall t \in \mathbb{T}_2 \qquad (2.2.28)$$

$$y_t \le y\_ub_t, \qquad \forall t \in \mathbb{T} \qquad (2.2.29)$$

$$y\_ub_t \le A_k \cdot \left(p_t - p_{wf}\right) + M \cdot \left(1 - x_{tk}\right), \quad \forall k \in \mathbb{K},\ t \in \mathbb{T} \qquad (2.2.30)$$

$$y\_ub_t \ge A_k \cdot \left(p_t - p_{wf}\right) - M \cdot \left(1 - x_{tk}\right), \quad \forall k \in \mathbb{K},\ t \in \mathbb{T} \qquad (2.2.31)$$

$$\sum_{k \in \mathbb{K}} x_{tk} = 1, \qquad \forall t \in \mathbb{T} \qquad (2.2.32)$$

$$p_{t+1} = p_t - B \cdot y_t \qquad \forall t \in \mathbb{T} \qquad (2.2.33)$$

$$p_t \ge \overline{p}_{\min} \qquad \forall t \in \mathbb{T} \qquad (2.2.34)$$

$$x_{tk} \in \{0, 1\}, \quad y_t \ge 0. \qquad \forall k \in \mathbb{K},\ t \in \mathbb{T}$$

The MILP model (2.2.26) ~ (2.2.34) attempts to find the optimal operating (drilling or shut-in) decisions x^* such that the resulting production plans y^* can recover the reservoir to the most. The optimal objective value gives us the maximal recovery ratio z^* / V_{OOIP}, where V_{OOIP} is estimated with formula (2.2.6).

The constraints (2.2.27) together with (2.2.28) and (2.2.29) impose technology and geological upper bounds to the annual production, respectively. The natural, or geological, production bound y_ub_i relies on the drilling decision through the mutually exclusive multiple choice constraints (2.2.30) ~ (2.2.32). The formula (2.2.32) says that at each decision point *i*, exactly one operating plan must be selected from the potential strategy set $\mathbb{K} = \{0, 1, \cdots, K\}$. If $x_{ik} = 1$, then *k* producing wells will be operated during the period $[t_i, t_{i+1})$. By (2.2.30) and (2.2.31), we can see that in this case, the geological upper bound will be forced to be exactly $y\_ub_i = A_k \cdot (p_i - p_{wf})$, which gives the maximal annual production under the driving force of the pressure difference $p_i - p_{wf}$ during the next year.

The constraint (2.2.33) models the linear relationship between the pressure drop and total production during a specific production period. The constraint (2.2.34) says the reservoir pressure should not drop below the minimal pressure level $\overline{p}_{\min}$, which is required to maintain economic production.

2.3 NUMERICAL EXAMPLE

We implement the optimal oilfield development model (2.2.26) ~ (2.2.34) in GAMS 22.5. The values of involved geological and operational parameters are listed in Table 2-3, borrowed from Hultzsch (2005). We solve the optimization model with CPLEX. The resulting optimal development plans and production profiles are summarized in the Figure 2-2 ~ Figure 2.4. The recovery rate is 14.26%.

Table 2-3: Reservoir parameters

Parameter	Symbol	Value	Parameter	Symbol	Value
Oil Permeability	k_o	373 mD	Skin factor	s	5
Porosity	ϕ	0.26	Reservoir pressure	$\overline{p}$	3991 psi
Oil viscosity	μ_o	10 cp	Well radius	r_w	0.5 ft
Oil saturation	S_o	0.72	Shape factor	C_A	31.62
Oil Compressibility	c_o	50 $Mpsi^{-1}$	Bottomhole pressure	p_{wf}	400 psi
Water Compress	c_w	1 $Mpsi^{-1}$	Max number of wells	$N_{\max}$	10
Formation Compressibility	c_f	0.05 $Mpsi^{-1}$	Oil formation volume factor	B_o	1.21 rb/STB
Drainage area	A	2084acres	Reservoir thickness	h	116 ft
Facility capacity	C	5.3 MMSTB	Expanded capacity	C_exp	1.7 MMSTB
Minimal pressure	$\overline{p}_{\min}$	420 psi	Project life	T	10 years

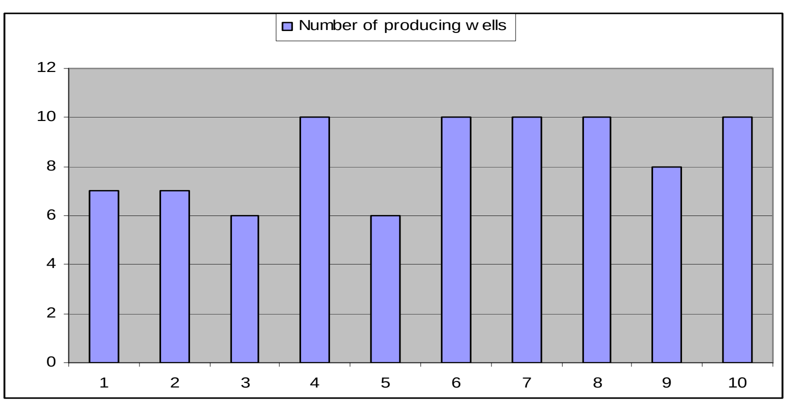


Figure 2-2: The optimal development plan for producing wells.

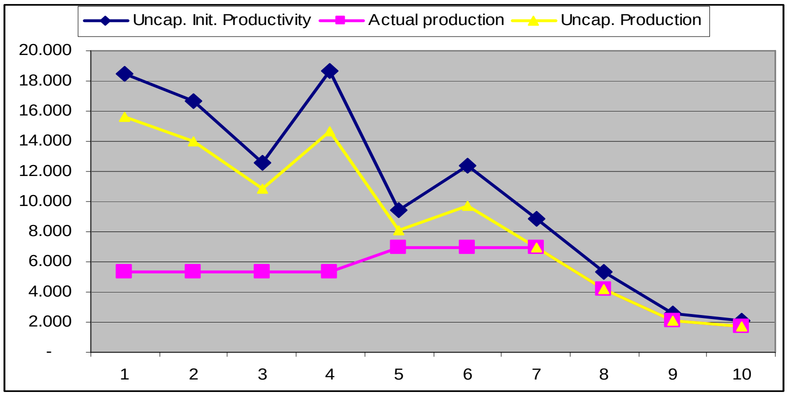


Figure 2-3: The reservoir production profiles correspond to the above optimal plan. The dark line is the envelope of initial production rate per period without capacity restriction. The yellow line is the annual production without capacity restriction. The pink line is the actual annual production curve. Notice, due to the capacity expansion at year 5, the pink line shows two plateau, the first four periods at level of 5.3 MMSTB and the fifth and the sixth periods at level of 7.0 MMSTB. After that, the reservoir production starts declining and the yellow line and pink line coincide.

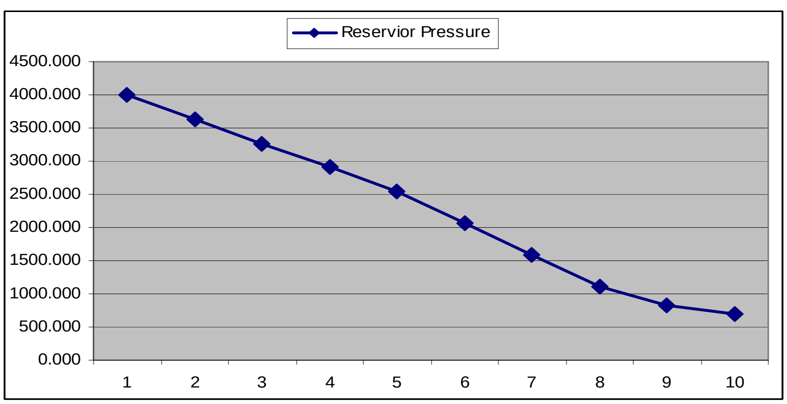


Figure 2-4: The reservoir pressure decline curve corresponds to the optimal development plan in Figure 2-2.

Chapter 3 Optimization Based Option Pricing Models

In this chapter, we consider the problem of prioritizing and developing explored but undeveloped oil fields over multiple time periods to maximize the total ENPV, subject to budget constraints and market uncertainty. We want to find out how the best project mix and the optimal drilling plans evolve as the budget level varies. For simplicity, we assume the uncertainties around the geological parameters of such oil fields have been significantly reduced. So the oil price dynamics plays the dominating role to determine the field development strategies. By assuming a complete market, we can apply the option pricing theories to evaluate individual fields and identify their optimal drilling plans. However, the traditional real options methods can not handle budget constraints and interactions among multiple projects. Therefore, new project valuation methods are needed. We will consider the similar but more general problems for unexplored oil fields with uncertain parameters in the following chapters.

Meier et al (2001) propose a dynamic capital budgeting model to deal with these problems. However, the model requires that the optimal exercising rules for all involved options be known beforehand either in closed-form or numerically. As a result, the model generates so-called "bang-bang" investment strategies in which either a project is excluded from the portfolio or it is allocated a fixed lump-sum of funds equal to the so-called striking price. Hence the model precludes the opportunity to partially develop an oil field even though the remaining fund is only sufficient for the partial development which would have added positive values to the portfolio. We extend this model in the later part of this chapter to allow the partial development as long as it does add positive value to the portfolio and it is optimal to do so.

In this chapter, we develop optimization based option pricing models and a dynamic capital budgeting model to tackle those problems. We take three steps to achieve those goals progressively. First, we develop optimization based option pricing models for both financial and real options. Second, we model the development of oil projects as real options and assess embedded investment opportunities by the optimization based option pricing models. Third, we develop a new real options based dynamic capital budgeting model, in which each oil project is modeled as a compound sequential option with our optimization pricing framework. The dynamic capital budgeting model can help to answer the foregoing questions, identify the best projects, and find out the optimal dynamic drilling schedules with a given initial budget. Our optimization based option pricing models and capital budgeting model have many advantages over traditional approaches, such as being easy to implement and flexible to extend. Particularly, our model can capture investment opportunities which would have been missed by other approaches, like Meier et al (2001).

This chapter is organized as follows. In §3.1, we give a literature review on optimization based option pricing models. We also present a numerical example to introduce how to model options pricing problems with optimization techniques. In §3.2, we model the option pricing problem as a linear program, based on the monotone and contraction properties of the corresponding dynamic programming mapping (Bertsekas, 2007). The LP based option pricing model is efficient to solve. However, it may be inconvenient to capture more complex exercising decisions, e.g. the integer drilling decisions in oil field development. Therefore, in §3.3, we start from the general asset pricing theory and develop an integer program (IP) option pricing model, in which exercise decisions are explicitly captured by binary variables. The MIP based option pricing model is essentially the stochastic programming version of the optimal stopping problem. In §3.4, we apply the MIP based option pricing framework to model the sequential development of oil fields and the capital budgeting problem. Finally, we conclude this chapter with an analysis to the numerical experiments of the capital budgeting model. The outcome shows the effectiveness of our methods.

3.1 INTRODUCTION

An American option gives its owner a right and not an obligation to buy or sell an underlying asset for a pre-specified price *K*, called the exercise price or strike price, at or before some maturity date *T*. The option owner should exercise the option optimally such that the future random payoffs received upon exercise fairly compensate the price he has paid now for owning the option.

In capital markets, investors are facing similar problems. They need to decide whether and when it is justified to commit a certain amount of money to carry out a project in return for a random stream of future payoffs. If the output of the project is tradable, as examples in Dixit and Pindyck (1994), we can treat the opportunity to invest the project as a call option. We can apply option pricing methods to identify the optimal investment strategies and discover the fair value to hold the project, which equals to the project's conventional ENPV plus a proper opportunity cost. We call such methods to assess real projects with option pricing methods as real options.

In complete markets, such optimal strategies are independent to investors' private time preference and risk attitude and can be assessed solely with public information and using the risk-neutral pricing method. This results in a unique project value. However, for incomplete markets where future payoffs depend on exogenous random events, the investors' individual time preference and risk attitude do matter to the optimal investment strategies and hence affect the project valuation. Consequently, different investors may value the same project differently and all of them should be justified as long as the investment strategies are optimal with respect to their beliefs and preferences. In this case, we can apply the expected utility theory to compare different projects.

No matter whether the markets are complete or not and the options are financial or real, in order to fairly valuate them, we need identify the optimal conditions to exercise the options or carry out the projects. Traditional methods indirectly deal with the underlying optimization problem and include converting it into a value-boundary problem with optimal control methods (Dixit and Pindyck, 1994), identifying a numerical optimal exercising boundary with simulation and regression (Long and Staff, 2001), and various lattice-like methods with dynamic programming.

In this chapter, we take a more straightforward approach to directly model the option exercising problem as optimization models. We will formulate the same option pricing problem in different forms, each with certain strength and weakness. All of them are linear models and some involve binary decision variables.

3.1.1 A literature review on option pricing optimization models

Existing linear programming (LP) based option pricing models fall into four categories: (1) formulating an equivalent LP for the linear complementary problem (LCP) resulting from the free boundary problem; (2) using LP to characterize the optimal stopping conditions (the time and the state to stop) which ensures the martingale property and maximizes the stopped reward; (3) using LP to model a dynamic replicating portfolio for the option and then price the option with the no-arbitrage argument; (4) using LP to directly model the pricing problem without constructing a replicating portfolio. The first two categories are deterministic linear programs and do not depend on scenarios.

The last two are stochastic models and involve distributions and scenarios. The first two types depend on the risk-neutral measure and the existence of a complete market, while the last two do not necessarily depend on the complete market assumption and may work under the incomplete markets setting as long as a scenario tree can be used. So far only Vanderbei and Pinar (2009) and our LP pricing model falls into the fourth category.

There are some relevant papers, like Ritchken (1985) and Christensen (2011), which aim to setup LPs to find good bounds for option prices.

Category one: equivalent LP for LCP

Dempster et al. (1998) and Dempster and Hutton (1999) consider the problem to derive an equivalent LP to solve the LCP of the optimal stopping problem. The LCP itself is a system of nonlinear equations. In order to formulate an equivalent LP whose solution (primal and dual) solves the LCP. They start from the linear complementary (LC) conditions and conduct a sequence of equivalent transformations from LC to variational inequality (VI) problem, then to least element (LE) problem, and finally to the equivalent LP. The equivalences are given by Cryer and Dempster (1980), based on the work of Cottle and Veinott (1972) and Cottle and Pang (1978). Other LP option pricing models developed by Dempster and his coauthors all follow the similar steps to derive an equivalent LP for the LCP. The equivalence conditions must be always checked before applying their ideas.

Category two: LP for martingale measures

Stockbridge (2004) develop a novel LP option pricing model to use linear constraints to characterize the martingale condition of an arbitrary set of bounded and continuous functions mapping from the time and state spaces to real. The martingale condition comes from the fact: the difference of a test function in a diffusion process over a short interval consists of two parts, the drift part and the diffusion part. Removing the drift part from the difference gives us a martingale process, whose (conditional) expectation is solely determined by the initial distribution. The author defines a measure on the stopping boundary, represented by the pair of the stopping time and stopping state, as decision variables then write down the martingale condition as linear constraints for each test function. The payoff obtained at stopping defines the objective function. He tries to find the optimal stopping boundary to maximize the stopped rewards. The work is based on the Manne (1960), Stockbridge (1990), and Cho and Stockbridge (2002). Similar ideas are shown in Helmes (2002a) and Helmes (2002b).

Christensen (2011) develops a new variant of Stockbridge's model to setup a tight upper bound for the option price. The model doesn't use any discretization scheme and doesn't use simulation as well. It leads to a semi-infinite linear programming (ILP). The problem is solved with a cutting plane approach which leads to an approximation.

The Stockbridge approach is completely different from ours and others. By the curse of dimensionality, this approach is limited to low dimensional problems.

Category three: LP for replication portfolio and non-arbitrage argument

The replicating portfolio approach to pricing options is based on the no-arbitrage argument. It represents the most popular and well-studied method in option pricing area. There are plenty of publications on using stochastic linear programming techniques to construct a replication portfolio. The well-known results are published in Huang and Litzenberger (1988), Korn and Korn (2000), and King (2002). More recent publications along this direction are Flam (2008) and Pinar and Camci (2009).

Since this is a relatively mature area and does not fit to our approach, so we skip the detailed reviews.

Category four: LP for the direct pricing problem

So far only Vanderbei and Pinar (2009) and our LP pricing model falls into the same category: use stochastic LP to model the pricing process. The "stochastics" is introduced by the direct uses of probabilities and scenarios. Dempster's LP option pricing models don't use distribution at all since they only work with the PDEs. Stockbridge et al.'s models take a different flavor by defining a special probability measures as the decision variables.

An additional review on integer programming (IP) based option pricing models

There exist two types of IP based option pricing models. Wang and de Neufville (2004) develop an optimization model to evaluate lattice based real options with MIP techniques. In their models, the exercising and holding decisions are explicitly captured by binary variables. However, the original models of Wang and de Neufville (2004) are nonlinear due to the product form of decision variables to compute the continuation value at each node on the scenario tree. In order to solve large problems, additional constraints are required to decouple the product forms with a special MIP technique called "discrete alternative" with big-Ms. The linearization (or decoupling) unavoidably increases the model size and the number of discrete variables and thus significantly deteriorates the numerical efficiency and stability.

After the completion of the dissertation, we become aware of the work of Pinar and Camci (2010). The idea of our IP based pricing model is basically the same as

Pinar and Camci's. However, our IP pricing model takes a more general asset pricing view and has been applied to assess complex real options. Our work in this direction is mainly motivated by Meier et al. (2001) and Longstaff and Schwartz (2001).

3.1.2 A numerical example: binomial option pricing

Let's start with a simple example, borrowed from Luenberger (1998), to use a binomial lattice model to price an American put. The underlying asset is a non-dividend paying stock. We assume the stock follows a Geometric Brownian Motion (GBM) process with current price $S_0 = \$62$, annual volatility $\sigma = 0.20$, strike price $K = \$60$, and expiration date $T = 5$ months. We also assume a fixed annual risk free rate $r_f = 0.10$.

We choose a step size of one month and evenly divide the time to expiration into five periods. After each time step, the stock price either goes up or down at proper rates and with corresponding transition probabilities. The time step (Δt), price change rates (u and d), and transition probabilities (p and q) are chosen such that the resulting lattice model converges to the continuous GBM process in the limit when the step size shrinks to zero. We will review how to determine those parameters in the next sections. For simplicity, we just list the results below.

The time step:	$\Delta t = 1/12$ (year).
The price going up rate:	$u = e^{\sigma\sqrt{\Delta t}} = 1.05943$.
The price going down rate:	$d = 1/u = 0.94390$.
Per step discount factor:	$R = 1 + r_f \cdot \Delta t = 1.00833$.
The risk-neutral probability of going up:	$p = (R-d)/(u-d) = 0.55770$.
The risk-neutral probability of dropping:	$q = 1 - p = 0.44230$.

Due to its recombining nature, it is convenient to represent a binomial lattice as an upper-triangular matrix in spreadsheet models and computer algorithms. The binomial lattice is shown in Figure 3-1. There are four values attached to each node, which are, from top down, the stock price, exercise value, continuation value, and option value.

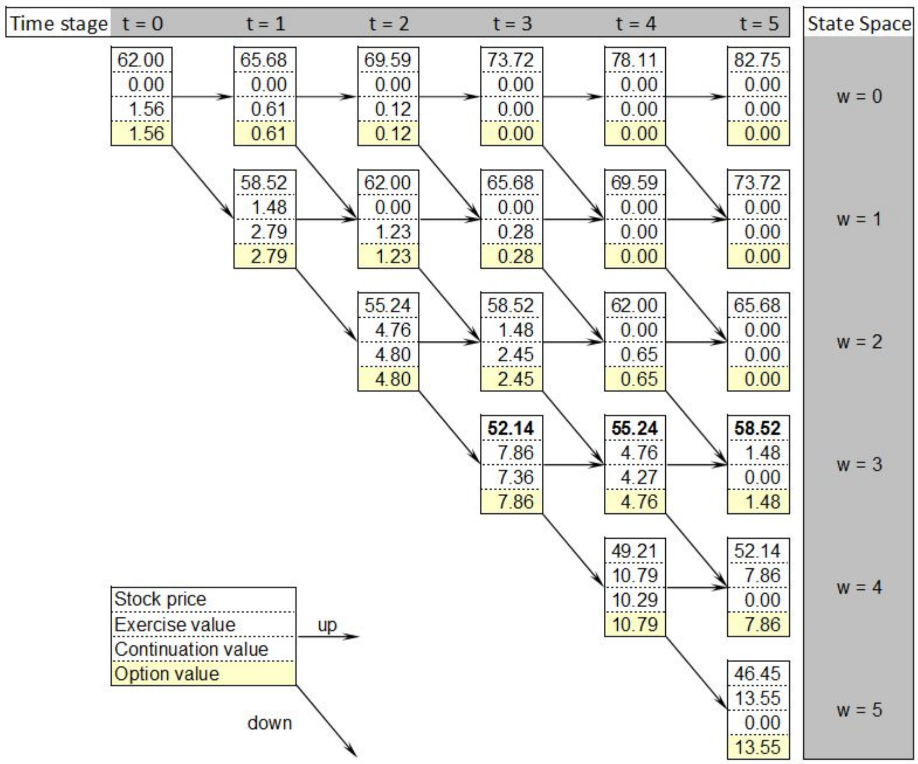


Figure 3-1 Binomial lattice model for the American put pricing (K=\$60,T=5)

We use a pair of coordinates (t, w) to denote a node, addressing its stage and state, respectively. We use $S_{t,w} = S(t,w)$ and $v_{t,w} = v(t,w)$ to denote the stock price and option value for the node (t,w). The time index t also counts the total number of jumps that have occurred. The state index w counts the number of downward jumps. For example, $S_{4,1} = S(4,1)$ is the stock price after four jumps, out of which exactly one jump is "down" and the other three are "ups." So $S_{4,1} = u^3 d \cdot S_0 = u^2 \cdot S_0 = S_{2,1} = 69.59$.

The binomial option pricing method is essentially an implementation of dynamic programming, which breaks the option value at each stage into two parts. The first part is the immediate profit received by executing the option right away. The second is the continuation value by holding the option alive to the next period in the hope for future optimal exercising.

We can calculate all those values in Figure 3-1 by rolling back the lattice with the following formulae,

$$v_{T,w} = \text{Max}\left\{K - S_{T,w}, 0\right\}, \quad \forall w = 0, \cdots, T,$$
$$v_{t,w} = \text{Max}\left\{K - S_{t,w}, \ \left(p \cdot v_{t+1,w} + q \cdot v_{t+1,w+1}\right) \big/ R\right\}, \quad \forall w = 0, \cdots, t, \quad \forall t = T-1, \cdots, 1, 0 .$$

One interesting observation is that each calculation of the above "rolling back" procedure can be formulated as a trivial linear programming (LP) problem,

$$v_{T,w} = \text{Min}\left\{y : y \geq K - S_{T,w}, \ y \geq 0\right\}, \quad \forall w = 0, \cdots, T,$$
$$v_{t,w} = \text{Min}\left\{y : y \geq K - S_{t,w}, \ y \geq \left(p \cdot v_{t+1,w} + q \cdot v_{t+1,w+1}\right) \big/ R\right\},$$
$$\forall w = 0, \cdots, t, \quad \forall t = T-1, \cdots, 1, 0 .$$

What is more interesting is that we can recursively combine smaller "naive" LPs into a bigger LP. Ultimately, we can price the option with only one big LP. To explain the idea clearly, let's consider a smaller option pricing problem to price $v_{3,3}$, the option value at node (t=3, w=3). We can calculate it by solving a "naive" LP,

$$v_{3,3} = \text{Min}\left\{y : y \geq K - S_{3,3}, \ y \geq \left(p \cdot v_{4,3} + q \cdot v_{4,4}\right) \big/ R\right\},$$

which relies on the optimal values of the following two "naive" LPs,

$$v_{4,3} = \text{Min}\left\{y : y \geq K - S_{4,3},\ y \geq \left(p \cdot v_{5,3} + q \cdot v_{5,4}\right)/R\right\}$$

and

$$v_{4,4} = \min\left\{y : y \geq K - S_{4,4},\ y \geq \left(p \cdot v_{5,4} + q \cdot v_{5,5}\right)/R\right\},$$

which further rely on the following three LPs,

$$v_{5,w} = \text{Min}\left\{y : y \geq K - S_{5,w},\ y \geq 0\right\},\ w = 3, 4, 5.$$

Instead of solving each of them separately, we can go backward recursively and combine smaller LPs into an equivalent bigger LP pricing problem. For example, to price $v_{4,3}$, we can pack the constraints of the two child-LPs associated with the two child nodes (5, 3) and (5, 4), of node (4, 3) to get an enlarged LP for $v_{4,3}$, which is

$$v_{4,3} = \text{Min}\left\{ y : \begin{array}{l} y \geq K - S_{4,3},\ y \geq \left(p \cdot v_{5,3} + q \cdot v_{5,4}\right)/R, \\ v_{5,3} \geq K - S_{5,3},\ v_{5,3} \geq 0, \\ v_{5,4} \geq K - S_{5,4},\ v_{5,4} \geq 0. \end{array} \right\}.$$

This problem doesn't rely on external values. Similarly, we can pack the node (4, 4)'s two child-LPs, $v_{5,4}$ and $v_{5,5}$, to get an enlarged LP for $v_{4,4}$. We can further pack the two enlarged LPs to their common parent node (3,3) to result in the enlarged LP,

$$\begin{aligned} v_{3,3} = \ & \underset{\{v_{w,t}\}}{\text{Min}} \quad y \\ \text{s.t.} \quad & y \geq K - S_{3,3}, \quad y \geq \left(pv_{4,3} + qv_{4,4}\right)/R, \\ & v_{4,3} \geq K - S_{4,3},\ v_{4,3} \geq \left(pv_{5,3} + qv_{5,4}\right)/R, \\ & v_{4,4} \geq K - S_{4,4},\ v_{4,4} \geq \left(pv_{5,4} + qv_{5,5}\right)/R, \end{aligned}$$

$$v_{5,3} \geq K - S_{5,3}\,,\quad v_{5,3} \geq 0$$

$$v_{5,4} \geq K - S_{5,4}\,,\quad v_{5,4} \geq 0$$

$$v_{5,5} \geq K - S_{5,5}\,,\quad v_{5,5} \geq 0$$

In this LP, two inequalities are associated to each node. The first inequality (e.g., $v_{4,3} \geq K - S_{4,3}$) bounds the nodal option value $v_{4,3}$ from below by its intrinsic value, i.e., the immediate exercise value. We call it the exercising inequality at node (4,3). The second inequality (e.g, $v_{4,3} \geq \left(pv_{5,3} + qv_{5,4} \right) / R$) of a non-leaf node bounds the option value from below by the continuation value. We call it the holding inequality at node (4,3). The second inequality of leaf nodes is simply non-negativity constraints.

After solving this linear program, the optimal solutions give the option value and the associated optimal exercise strategy at each node. The strategies can be recovered by inspecting the inequalities' binding condition. Extensive numerical experiments using GAMS/CPLEX suggest that most time the LP pricing problem can be readily solved during the pre-process step. In the following development, we can see this LP option pricing model is essentially a finite horizon generalization of the approach to use linear programming to solve an infinite horizon dynamic programming problem.

We finish this numerical example by presenting the full LP option pricing model as below. More general discussions about it will be left for the next section.

$$\begin{array}{lll}
\min\limits_{\{v_{t,w}\}} & v_{0,0} & \\
\text{s.t.} & v_{t,w} \geq K - S_{t,w}\,, & \forall w = 0,\cdots,t\,,\ \ \forall t = T-1,\cdots,1,0\,, \\
& v_{t,w} \geq \left(pv_{t+1,w} + qv_{t+1,w+1} \right) / R\,, & \forall w = 0,\cdots,t\,,\ \ \forall t = T-1,\cdots,1,0\,, \\
& v_{T,w} \geq K - S_{T,w}\,, & \forall w = 0,\cdots,T\,, \\
& v_{T,w} \geq 0\,, & \forall w = 0,\cdots,T\,.
\end{array}$$

3.1.3 Fundamental price process

In the remaining discussions of this chapter, we consider the evaluation of risky projects. We assume the output of the completed project is a tradable asset, whose spot price $S(t)$ follows an GBM process under the real world probability measure $\mathbb{P}(\cdot)$,

$$dS(t) = \alpha S(t)dt + \sigma S(t)dW(t),$$

where α is the rate of price appreciation, σ is the volatility, and $W(t)$ is normally distributed with mean of 0 and variance of t. Moreover, $\{W(t), t \geq 0\}$ forms an Brownian motion under $\mathbb{P}(\cdot)$ and $dW(t) \equiv W(t+dt) - W(t) \sim N(0, dt)$ is independent of $W(t)$.

Since the asset is tradable, the capital asset pricing model (CAPM) theory[3] claims that any investor would require a total rate of return μ ($\mu \geq \alpha$) to hold it. For storable consumption commodities like crude oil, the total rate of return μ breaks down into two components, the tangible price appreciation α and the intangible convenience yield δ (net the cost of carry), which is received by physically owning the asset. As a result, $\mu \equiv \alpha + \delta$. This equality represents an equilibrium condition and is always true. Generally, we assume $\delta = \mu - \alpha > 0$ to ensure that investors do have incentives to hold the commodity. Equivalently, we can rewrite the asset price process as

---

[3] The CAPM theory implies that the expected rate of return of an asset S is $\mu = r + \phi\sigma\rho_{sm}$, where r is the risk free rate, σ the asset volatility, ρ_{sm} the correlation between the asset and the market, $\phi = (r_m - r)/\sigma_m$ the market price of risk, r_m the market rate of return, and σ_m the market volatility.

$$dS(t) = (\mu - \delta) S(t) dt + \sigma S(t) dW(t).$$

We usually price derivatives under a risk neutral measure $\tilde{\mathbb{P}}(\cdot)$. Let *r* denote the risk free rate. By Girsanov Theorem (Shreve, 2004), the shifted process $\tilde{W}(t) = \frac{\mu - r}{\sigma} t + W(t)$ is a Brownian motion process under the risk-neutral probability measure $\tilde{\mathbb{P}}(\cdot)$. So the price process under $\tilde{\mathbb{P}}(\cdot)$ is expressed as a new GBM process

$$dS(t) = (r - \delta) S(t) dt + \sigma S(t) d\tilde{W}(t), \quad t \geq 0, \tag{3.1.1}$$

where *r, δ,* and *σ* all are nonnegative annualized constants with $r \geq \delta$. Since $d\tilde{W}(t)$ follows a normal distribution $N(0, dt)$, the price percentage change, $dS(t)/S(t)$, follows a normal distribution $N\left((r - \delta) dt, \sigma^2 dt\right)$. The following is the unique solution to (3.1.1). It says the asset price at any time *t* follows a lognormal distribution,

$$S(t) = s e^{\left(r - \delta - \frac{1}{2}\sigma^2\right) t + \sigma \tilde{W}(t)}. \tag{3.1.2}$$

In the following discussions, we will frequently use both lattices (binomial and trinomial) and grids (finite differences) to approximate the prices evolution. In those cases, it is more convenient to work with the logged price. Let $X(t) = \ln S(t)$ and $x = \ln s$ to denote the logged price at time *t* and the initial logged price, respectively. So $S(t) \equiv e^{X(t)}$. Let's define $\upsilon = \left(r - \delta - \frac{1}{2}\sigma^2\right)$. By the Ito-Doeblin formula, the logged price follows the Ito process as below,

$$dX(t) = \upsilon dt + \sigma d\tilde{W}(t), \quad t \geq 0, \tag{3.1.3}$$

$$X(t) = X(0) + \int_0^t \upsilon du + \int_0^t \sigma d\tilde{W}(u) = x + \upsilon t + \sigma \tilde{W}(t), \quad t \geq 0. \tag{3.1.4}$$

The main advantage to work with the logged price instead of the original price becomes obvious if we compare (3.1.1) with (3.1.3). The drift and the diffusion parameters in (3.1.3) are both constant and independent of the current state variable $X(t)$. This property facilitates the construction of the succeeding lattice model and finite difference model (Clewlow and Strickland, 1998).

3.2 Optimization Option pricing Models: Linear Programming

In this section, we develop a sequence of optimization models to price financial options. First, we formulate the perpetual American put pricing problem as a LP problem. We show this method is essentially the same as solving an infinite horizon dynamic programming with linear programming (Bertsekas, 2007). Second, we generalize the LP perpetual option pricing model to price general options with a finite expiration date (like the example in §3.1.1). Consequently, we can extend the LP approach to solve infinite horizon dynamic programming problems to finite horizon cases. The LP option pricing model is efficient. The optimal exercise-continuation boundary can be easily obtained from the LP solutions. However, sometimes, it may be more convenient or preferable to directly address those "go-no go" decisions with binary variables. Finally, therefore, we introduce two additional mixed integer programming (MIP) option pricing models. Those optimization models are useful in evaluating real options. When they work together, they can model and solve more interesting and complicated project valuation and investment problems under market uncertainty.

3.2.1 Pricing perpetual American put options

When the underlying price follows the GBM process, we can price a perpetual American option in closed-form. At any time, the perpetual American option always has an infinite time-to-expiration. So its value does not rely on the time index but only on the present stock price. Consider a perpetual American put. Given the stock price *s*, the put value $v(s)$ is determined by the infinite horizon optimal stopping problem

$$v(s) = \underset{\tau \in \mathcal{T}}{\text{Max}} \; \tilde{\mathbb{E}}\left[e^{-r\tau} \left(K - S(\tau)\right)^{+} \middle| S(0) = s \right], \tag{3.2.1}$$

where $(x)^{+} = \max\{0, x\}$ for any real *x*, $S(\tau)$ follows the GBM process (3.1.1) with zero dividends (i.e., $\delta = 0$), the expectation is under the risk neutral measure, and the set $\mathcal{T}$ consists of all valid exercising times. For perpetual options, $\mathcal{T}$ contains the whole time axis plus the infinite element, i.e., $\mathcal{T} = \{t \geq 0\} \cup \{\infty\}$, where we employ the convention that the decision $\tau = \infty$ implies that the option is never exercised. Any element τ in the set $\mathcal{T}$ must be a stopping time, which means τ must be chosen only based on the observations up to and include time τ, $\{S(t), 0 \leq t \leq \tau\}$ [4], and should not rely on any information revealed in the future. In operations research society, people more frequently refer to the stopping time restriction as the "non-anticipativity" condition, meaning the present decisions can not be made on future anticipated events.

The problem (3.2.1) models the option owner's problem, which is, given any arbitrary current price *s*, how to choose the optimal exercise time τ such that the expected

[4] For GBM processes with constant parameters, the processes $\{S(u)\}_{0 \leq u \leq t}$, $\{\tilde{W}(u)\}_{0 \leq u \leq t}$ under $\tilde{\mathbb{P}}(\cdot)$, and $\{W(u)\}_{0 \leq u \leq t}$ under $\mathbb{P}(\cdot)$ convey the same information at *t*. The information can be described by a σ-algebra and denoted as $\mathcal{F}_t = \sigma(W(t))$. The infinite sequence $\{\mathcal{F}_t\}_{t \geq 0}$ forms a filtration and represents how information evolves over time. In stochastic programming, the filtration is usually approximated by a scenario tree and the stopping time restriction is normally enforced by the non-anticipativity constraints.

present value of the exercising payoff is maximized. Again, the expectation $\tilde{\mathbb{E}}[.]$ is taken under the risk-neutral probability measure $\tilde{\mathbb{P}}(\cdot)$ conditional on current information.

Analytical solution

There are different ways to find out the closed-form formulae for the price of perpetual American options. Shreve (2004) first uses the concept of "hitting time" and a Laplace transformation technique to identify the optimal exercising boundary L^* and the expected discount factor for the time when the underlying price first hits the boundary. With these results, he gets the option price formulae in both sides of the boundary. Dixit and Pindyck (1994) first argue that there is an exercise boundary dividing the pricing problem into two separate regions, the continuation region and the stopping region. The imaginary boundary is called the free boundary since it is unknown and must be obtained as part of the solution of the pricing problem. The authors separately apply two different approaches (dynamic programming and contingent claim analysis) to identify the same kind of ordinary differential equation (ODE) that the option price must conform in the continuation region. In either approach, they essentially solve a so-called "free boundary" problem. The detailed procedures can be found in Shreve (2004, p.345-352) and Dixit and Pindyck (1994, Chapter 5). We simply put the results below as a reference.

From the pricing problem (3.2.1), it is natural to guess that the optimal exercising policy L^* should not depend on time due to the infinite time-to-expire and, intuitively, L^* should be smaller than *K*. Whenever the stock price hits or falls below L^*, the put is exercised. The respective optimal exercising boundary and the option price are,

$$L^* = \frac{2r}{2r+\sigma^2}K \,, \tag{3.2.2}$$

$$v(s) = \begin{cases} K - s & , \ 0 \le s \le L^* \\ \left(K - L^*\right)\left(\dfrac{s}{L^*}\right)^{-\frac{2r}{\sigma^2}} & , \ s \ge L^* \end{cases} . \tag{3.2.3}$$

We will use the analytical results as the benchmark to verify the correctness and effectiveness of our optimization-based option pricing model.

Linear programming formulation

For any logged price $x = \ln(s)$ and $x \in (-\infty, \infty)$, the option price is

$$f(x) = \underset{\tau \ge 0 \text{ and } \tau \in \mathcal{T}}{\text{Max}} \ \tilde{\mathbb{E}}\left[e^{-r\tau}\left(K - e^{X(\tau)}\right)^+ \middle| X(0) = x\right], \tag{3.2.4}$$

where $X(\tau)$ is driven by (3.1.3). Since the log function is strictly monotone and therefore a one-to-one mapping, so $v(s) \equiv f(x)$ for all strictly positive s.

If we treat the option price $f(\cdot)$ as the so-called "cost-to-go" function, we can express the pricing problem (3.2.4) as dynamic programming with binary decisions,

$$f(x) = \text{Max}\left\{\left(K - e^x\right)^+, \ e^{-rdt}\tilde{\mathbb{E}}\left[f(x+dx)\middle| X(0) = x\right]\right\}, x \in (-\infty, \infty). \tag{3.2.5}$$

The equation (3.2.5) is essentially the Bellman's equation for the perpetual American put pricing problem. It says that the price of the option is equal to either its intrinsic value $\left(K - e^x\right)^+$ if it is exercised immediately or the expected present value of holding it alive over a short interval dt, during which the logged price evolves from x to $x + dx$.

Since the intrinsic value of the put is bounded from above by the strike price K and the discount factor is strictly positive, by properties of dynamic programming (Bertsekas, 2007), there exists a unique functional $f(\cdot)$ satisfying the Bellman equation (3.2.5). By the monotonicity of the Bellman equation, the unique solution $f(\cdot)$ to the Bellman equation is a contraction mapping which must be the solution of the following linear programming problem,

$$\begin{aligned}
&\underset{f(\cdot)}{\text{Min}} \int_{-\infty}^{+\infty} c(x)\cdot f(x)\,dx \\
&s.t.\;\; f(x)\geq K-e^{x},\;\; x\in(-\infty,\infty), \\
&\qquad f(x)\geq e^{-rdt}\cdot\tilde{\mathbb{E}}\left[f(x+dx)\middle|X(0)=x\right],\;\; x\in(-\infty,\infty),
\end{aligned}$$

where we have dropped the + sign of the intrinsic value for simplicity and without changing the problem, $c(x)$ can be any finitely bounded function such that $c(x)>0$, $\forall x\in(-\infty,\infty)$, and x denotes the logged price. For simplicity, we choose $c(x)\equiv 1$, $\forall x\in(-\infty,\infty)$. In fact, any proper choice of $c(x)$ would lead to the same solution. It is easy to justify that the optimal solution to this infinite linear program satisfies (3.2.5) by inspection. We also can justify it by the optimal solution of its dual problem, which is easy to formulate and solved with inspection as well. The objective functions of the primal and dual problems coincide at optimality.

However, the above LP is defined on an unbounded domain $(-\infty,\infty)$ and with infinite number of constraints. Fortunately, in practice, the stock price reflects the equilibrium between supply and demand and should not grow wild. We can truncate and discretize its logged domain $(-\infty,\infty)$ to ensure the resulting LP can be solved practically. Let $\mathbb{X}$ denote such a discretization scheme with a finite number $|\mathbb{X}|=N$ of points in $(-\infty,\infty)$. We can write the approximate pricing problem (3.2.5) as

$$\underset{f(\cdot)}{\text{Min}} \sum_{x \in \mathbb{X}} f(x) \tag{3.2.6}$$

$$s.t. \quad f(x) \geq K - e^x, \quad x \in \mathbb{X}, \tag{3.2.7}$$

$$f(x) \geq e^{-rdt} \cdot \tilde{\mathbb{E}}\left[f(y) \middle| X(0) = x \right], \quad x \in \mathbb{X}, \tag{3.2.8}$$

where state variable $y \in \mathbb{X}$ in the right-hand-side of (3.2.8) denotes the succeeding state transited from state *x* after a short time interval *dt*. To deal with the conditional expectation, we introduce two popular approximation approaches, the finite difference and the trinomial (or binomial) lattice.

It has been shown that the long-run stationary distribution of a bounded GBM process with reflective (instead of absorptive) boundaries converges to a simple exponential distribution (Dixit and Pindyck, 1994).

Linear programming: Finite difference

In (3.2.8), the state variation $dx = y - x$ is governed by the SDE (3.1.3),. Notice $e^{-rdt} \approx (1 + rdt)^{-1}$ as long as the interval *dt* is short enough. After we plug them into (3.2.8) and properly rearrange the terms on both sides, we get

$$\begin{aligned} rf(x)dt &\geq \tilde{\mathbb{E}}\left[f(x+dx) - f(x) \middle| x \right] = \tilde{\mathbb{E}}\left[df(x) \middle| x \right] \\ &= \tilde{\mathbb{E}}\left[\left(\frac{1}{2}\sigma^2 f''(x) + \upsilon f'(x) \right) dt + \sigma f'(x) d\tilde{W}(x) \middle| x \right] \\ &= \left(\frac{1}{2}\sigma^2 f''(x) + \upsilon f'(x) \right) dt, \end{aligned}$$

where the second line is obtained by applying the Ito-Doeblin's formula to $df(x)$. Eliminating *dt* on both sides of the last line gives an inequality equivalent to (3.2.8) as of

$$rf(x) \geq \frac{1}{2}\sigma^2 f''(x) + \upsilon f'(x), \quad x \in \mathrm{X}. \tag{3.2.9}$$

However (3.2.9) involves the first and the second derivatives of f. A natural way to approximate the derivatives is to use finite difference. So we should choose such a discretization scheme $\mathbb{X}$ that it facilitates the application of the finite difference. The simplest way is to choose a fixed step Δ_x and evenly pick *N* consecutive points, $x_1 < x_2 < \cdots < x_N$, such that $x_{i+1} - x_i = \Delta_x$ for any *i*. We also need ensure that the probability for points falling out the range is negligible. We then can apply central differences to the derivatives of *f* on the RHS of (3.2.9). The resulting LP is

$$\underset{\{f_i\}}{\mathrm{Min}} \ \sum_{i=1}^{N} f_i \tag{3.2.10}$$

$$s.t. \quad f_i \geq K - S_i, \ 1 \leq i \leq N, \tag{3.2.11}$$

$$rf_i \geq \frac{1}{2}\sigma^2 \frac{f_{i+1} - 2f_i + f_{i-1}}{{\Delta_x}^2} + \upsilon \frac{f_{i+1} - f_{i-1}}{2\Delta_x}, \ 2 \leq i \leq N-1, \tag{3.2.12}$$

$$f_1 = f_2 + \left(e^{x_2} - e^{x_1}\right), \tag{3.2.13}$$

$$f_N = f_{N-1}. \tag{3.2.14}$$

Notice, the inequality (3.2.12) cannot handle the bottom and top boundary points {*1, N*} since this would require f_0 and f_{N+1}, which we don't have. As a consequence, to deal with the bottom and top points, we introduce the equalities (3.2.13) and (3.2.14). The neighborhood of the state 1 corresponds to small prices such that $s_1 = e^{x_1} << K$ and $s_2 = e^{x_2} = e^{x_1 + \Delta_x} << K$, where the put will be exercised right away. So we have $\frac{df}{dx}\Big|_{x_1} = -1$, so $\frac{f_2 - f_1}{e^{x_2} - e^{x_1}} = -1$, so we get (3.2.13). However, things get involved at the top boundary where $s_N = e^{x_N} >> K$ and *f* reflects the holding value which is hard to get.

For simplicity, we use a crude approximation $f_N = f_{N-1}$ since the option value changes continuously and will not deviate too much within a small range.

So far we have identified all constraints required to characterize the option value on the grid. One interesting result of the LP model (3.2.10) ~ (3.2.14) is the time index indeed does not come into play and we don't need do anything to discretize the time axis. This is slightly different from the next trinomial tree approach, which requires a proper time step to discount the continuation value over the short period *dt*.

Linear programming: Trinomial tree

Another approach to approximate the conditional expectation in (3.2.8), $\tilde{\mathbb{E}}\left[f\left(x+dx\right)\middle|x\right]$, is to use lattice models, either the binomial lattice or the trinomial lattice. We use the trinomial lattice as an example to achieve the approximation since the trinomial lattice is equivalent to the (explicit) finite difference when the time and price steps are chosen properly.

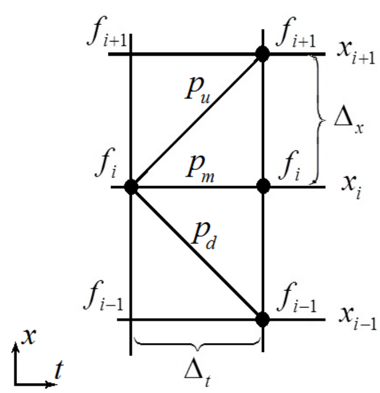


Figure 3-2: The discretization scheme by the trinomial lattice.

We assume the asset price x may transit to anyone of three possibilities after a fixed short interval Δ_t: $x+\Delta_x$, x, and $x-\Delta_x$, which are denoted as up, middle, and down, respectively. The corresponding one-step stationary transition probabilities are p_u, p_m, and p_d.

The time step Δ_t and the (logged) price step Δ_x can not be chosen independently to ensure numerical stability and convergence. A good choice is $\Delta_x = \sigma\sqrt{3\Delta_t}$, as suggested by Clewlow and Strickland (1998). To determine the transitional probabilities, we write (3.1.3) in the difference approximation form as

$$\Delta x \approx \upsilon\Delta_t + \sigma\sqrt{\Delta_t} \cdot \tilde{W}(1), \tag{3.2.15}$$

where $\tilde{W}(1)$ follows a standard normal distribution. (Notice: the above Δx denotes the variation of the logged price over a short time Δ_t, i.e., $\Delta x = X(t+\Delta_t) - X(t)$. It is not the chosen fixed space step Δ_x.) Once the time and space steps are chosen, we can calibrate the transitional probabilities by matching them to the moments of Δx implied by (3.2.15), which lead to three equalities

$$\tilde{\mathbb{E}}\left[\Delta x \middle| X(0) = x\right] = p_u \cdot \Delta_x + p_m \cdot 0 + p_d \cdot (-\Delta_x) = \upsilon\Delta_t,$$
$$\tilde{\mathbb{E}}\left[\Delta x^2 \middle| X(0) = x\right] = p_u \cdot \Delta_x{}^2 + p_m \cdot 0 + p_d \cdot (-\Delta_x)^2 = \sigma^2\Delta_t + \upsilon^2\Delta_t{}^2,$$
$$p_u + p_m + p_d = 1.$$

Solving above equations gives

$$p_u = \frac{1}{2}\left(\frac{\sigma^2\Delta_t + \upsilon^2\Delta_t{}^2}{\Delta_x{}^2} + \frac{\upsilon\Delta_t}{\Delta_x}\right),\ \ p_m = 1 - \frac{\sigma^2\Delta_t + \upsilon^2\Delta_t{}^2}{\Delta_x{}^2},\ p_d = \frac{1}{2}\left(\frac{\sigma^2\Delta_t + \upsilon^2\Delta_t{}^2}{\Delta_x{}^2} - \frac{\upsilon\Delta_t}{\Delta_x}\right).$$

Therefore, the random variation Δx may take three values Δ_x, 0, and -Δ_x, with probabilities p_u, p_m, and p_d, respectively. Once the five parameters (Δ_x, Δ_t, p_u,

p_m, and p_d) are determined, we can estimate the conditional expectation of the continuation value by holding the option over a short period Δ_t as

$$\tilde{\mathbb{E}}\left[f\left(x+\Delta x\right) \middle| X\left(0\right)=x \right] \approx p_u f\left(x+\Delta_x\right)+p_m f\left(x\right)+p_d f\left(x-\Delta_x\right).$$

For simplicity, we choose the same discretization scheme as the previous finite difference model and pick the time step accordingly. The resulting trinomial lattice based LP pricing problem is

$$\underset{\{f_i\}}{\text{Min}} \ \sum\nolimits_{i=0}^{N} f_i$$

$$s.t. \quad f_i \geq K - S_i, \ 1 \leq i \leq N, \tag{3.2.16}$$

$$f_i \geq e^{-r\Delta_t}\left(p_u f_{i+1} + p_m f_i + p_d f_{i-1}\right), \ 2 \leq i \leq N-1, \tag{3.2.17}$$

$$f_1 = f_2 + \left(e^{x_2} - e^{x_1}\right), \tag{3.2.18}$$

$$f_N = f_{N-1}. \tag{3.2.19}$$

Numerical results

Our numerical experiments show that both the truncation and discretization schemes impact the solution accuracy. Particularly, when we fix one factor and adjust the other to try different configurations, we found that the numerical accuracy is more sensitive to the truncation scheme. Even for a rather crude discretization, we may achieve good result if the truncation range is not too narrow. This is reasonable since the truncation errors propagated from the truncation boundaries can be diminished by push the boundaries far away from the interested price range.

We price a perpetual American put written on a stock whose volatility is 20% and strike price is \$100. The annual risk-free interest rate is 6% and the stock pays no dividends. The truncated price range is selected to be $[\$0.01, \$1000]$, which is sufficient to cover realistic values the stock price may take. Notice, we require the lower bound to be strictly positive. We implemented both the finite difference (FD) and the lattice (LAT) based pricing models in GAMS and use CPLEX to solve them.

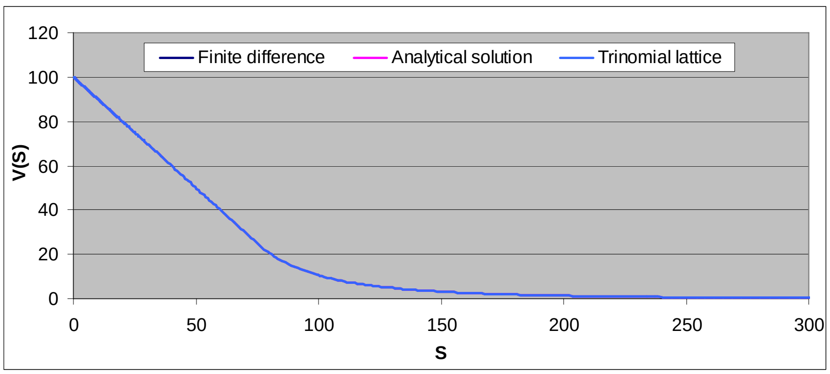


Figure 3-3: The option price V(S) as a function of the underlying asset price S.

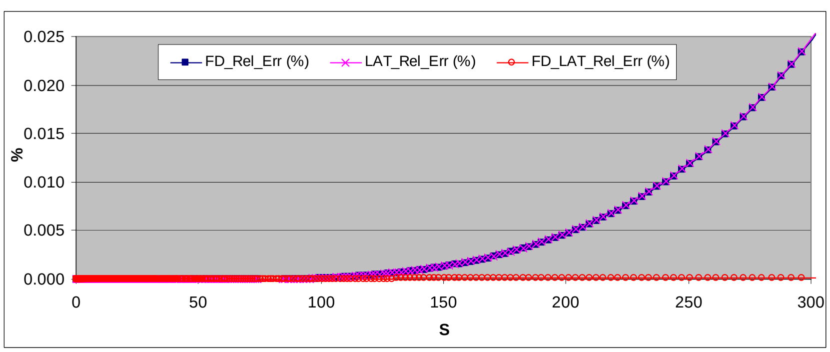


Figure 3-4: The relative errors of the solutions produced by the three methods.

There are different strategies to improve the numerical accuracy, stability, and efficiency of both finite difference and lattice methods. Many of them (such as Crank-Nicolson) can be implemented in our optimization models as well.

3.2.2 Pricing finite-expiry American options

Perpetual options do not exist in the real world. All traded options have finite expiration dates. Using a scenario tree to approximate the price process, we can generalize the previous LP model to price options with finite expiration dates. Again, we will formulate the pricing problem of American options with a finite expiry as an optimization problem (3.2.20) as follows.

Suppose the logged asset price at time *t* is $x(t)=x$, $0 \le t \le T$ and $x \in \mathbb{R}$. Let $\tau \in [t,T] \cup \{\infty\}$ denote an exercise time (or stopping time) and $h(x(\tau))$ be the payoff value upon exercising, where *T* is the expiration date and $h:\mathbb{R} \to \mathbb{R}_+$ is a deterministic payoff function. The time-*t* option value $v(t,x)$ corresponding to the deterministic exercising policy τ is given by the expected present value of the random payoffs,

$$v(t,x) = \tilde{\mathbb{E}}^{t,x}\left[e^{-r(\tau-t)}h(x(\tau))\right],$$

where the risk-neutral conditional expection $\tilde{\mathbb{E}}^{t,x}[\,\cdot\,] = \tilde{\mathbb{E}}\left[\,\cdot\,\middle|\,x(t)=x\right]$. The option pricing problem is hence to ask how to choose an optimal exercise policy τ^* such that the option value is maximized, i.e.,

$$v^*(t,x) = \tilde{\mathbb{E}}^{t,x}\left[e^{-r(\tau^*-t)}h(x(\tau^*))\right] \ge \tilde{\mathbb{E}}^{t,x}\left[e^{-r(\tau-t)}h(x(\tau))\right], \text{ for all policy } \tau.$$

Mathematically, we can represent this prcing problem as an optimization problem

$$v^*(t,x) = \underset{\tau \in \mathcal{T}_{[t,T]\cup\{\infty\}}}{\mathrm{Max}} \tilde{\mathbb{E}}^{t,x}\left[e^{-r(\tau-t)}h(x(\tau))\right], \ \ 0 \le t \le T, x \in \mathbb{R}, \tag{3.2.20}$$

where $\mathcal{T}_{[t,T]\cup\{\infty\}}$ requies all exercise times to be stopping times and the logged price $x(t)$ follows the SDE (3.1.3). Particularly, $h(x) = (e^x - K)^+$ for American calls, and $h(x) = (K - e^x)^+$ for American puts. We can drop the "+" sign of the non-negative operator "$(\cdot)^+$" as explained previously. In fact, our succeeding model allows more general stochastic price processes as long as they can be approximated by scenario trees.

The problem (3.2.20) restricts that the exercising time τ must chosen prior to the expiration of the option such that the expected present value of received random payoffs is the largest. By convention, decision $\tau = +\infty$ implies the option is abandoned due to the resulting infinitesimal discount factor. Since the exercise decision is binary, we can reformulate (3.2.20) as a optimal stopping problem as below,

$$v^*(t,x) = \mathrm{Max}\left\{ h(x(t)), \frac{1}{1+r\,dt}\tilde{\mathbb{E}}^{t,x}[v^*(t+dt, x+dx)] \right\},$$

$$0 \le t \le T, x \in \mathbb{R}. \tag{3.2.21}$$

This optimization problem (3.2.21) is called Bellman's equality. It expresses the Bellman's Principle of Optimality (Dixit and Pindyck, 1994, and Bertsekas, 2007) that the option value equals either the immediate payoff $h(x)$ if we exercise the option right away or the continuation value otherwise. Furthermore, the continuation value equals the expected present value of the option price $v^*(t+dt, x+dx)$ if we hold the option alive over a short period dt. The option price $v^*(t+dt, x+dx)$ is determined by optimally exercising the option starting from $t+dt$ and onward up to the expiration date. The vlaue $v^*(t+dt, x+dx)$ is recursively defined by (3.2.21).

Before we continue to generalize the LP pricing model to the finite expiration cases, we need introduce the concept of scenario tree, which is required to express how the underlying price evolves and how we model contingent exercising strategies.

Scenario tree

An asset's price evolves continuously over time. To model and price its derivatives with stochastic programming methods like the LP based perpetual options pricing model, we need approximate the asset price process with a discrete one. It is convenient to organize the resulting discrete process in a scenario tree. On this tree, a leaf node determines a unique sample path starting from the root node at time *0* and ending at the leaf node at time *T*. Each path (or each leaf node) represents one complete revelation of the state of nature. So the complete set of leaf nodes $\boldsymbol{\Omega}$ defines the full state space. Any intermediate node only represents partial revealed information.

The structure of a scenario tree can be defined by its node set $\boldsymbol{N}$, pairwise precedence-successor relationship defined on $\boldsymbol{N}$, associated transition probabilities, and nodal state values. The nodes are uniquely indexed, with the root node being indexed by *0*. We use a subset N_t of $\boldsymbol{N}$ to denote the nodes in stage *t*, for $t = 0, 1, \cdots, T$. Clearly, N_0 only contains the root node *0* and N_T contains all leaf nodes. As a result, the family of node sets $\{N_t, t = 0, \cdots, T\}$ defines a discrete filtration.

To organize the elements in the filtration as a scenario tree, we define the transitional relationship between pairs of nodes in any two consecutive node sets. For any node $n \in N_t$, $1 \le t \le T$, it has a unique parent node $a(n) \in N_{t-1}$. For any node $n \in N_t$, $0 \le t \le T-1$, it has a set of child nodes $C(n) \subseteq N_{t+1}$. We use $p_{n,m}$ to denote the transition probability from node n to node m such that $m \in C(n)$ and then $\sum_{m \in C(n)} p_{n,m} = 1$ for any non-leaf node $n \notin N_T$. If we use p_n to denote the (non-conditional) probability of node n, then we can compute it with the recursive relation $p_n = p_{a(n)} \cdot p_{a(n),n}$ and the root node probability $p_0 = 1$. The transition probability is essentially the conditional probability such that $p_{a(n),n}$ = $P\{n_{t+1} = n \mid n_t = a(n)\}$ = $p_n / p_{a(n)}$ for any non-root node n, or $p_{n,m}$ = $P\{n_{t+1} = m \mid n_t = n\}$ = p_m / p_n for any $m \in C(n)$ and non-leaf node n. With the conditional probability, we can perform the conditional expectation calculation which is required to compute continuation values in lattice based option pricing methods.

The Figure 3-5 below provides a simple example for a (symmetric) scenario tree with two periods and seven nodes.

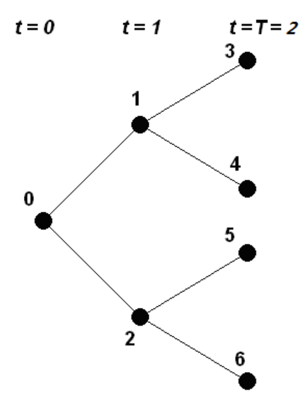


Figure 3-5: A scenario tree with two periods or three stages and totally seven nodes. For example, $N = \{0, \cdots, 6\}$, $N_0 = \{0\}$, $N_1 = \{1, 2\}$, $N_2 = \{3, \cdots, 6\}$, $C(1) = \{3, 4\}$, $a(1) = 0$, and $a(5) = 2$.

Optimization pricing models

We can formulate the optimal stopping problem (3.2.21) as an equivalent LP problem as below, for $0 \le t \le T$, $x \in \mathbb{R}$,

$$v^*(t,x) = \underset{v(t,x)}{\text{Min}} \ v(t,x) \tag{3.2.22}$$

$$s.t. \quad v(t,x) \geq h(x(t)), \tag{3.2.23}$$

$$v(t,x) \geq \frac{1}{1+r\,dt} \tilde{\mathbb{E}}^{t,x}[v^*(t+dt, x+dx)]. \tag{3.2.24}$$

The comparisons between the two optimal stopping problems (3.2.5) and (3.2.21) and their respective LP formulations (3.2.6)~(3.2.8) and (3.2.22)~(3.2.24) show that they are essentially the same idea. The feasibility of this formulation requires the option value at time t and in state x must be no less than both the immediate exercise value and the continuation value if the option is held alive over a short period of length dt. The optimality rules out any redundancy and requires the option value exactly be the larger of the exercise value and continuation value, i.e., at least one of the constraints must be binding at optimality.

As long as we can approximate the underlying stochastic process $\{x(t), 0 \leq t \leq T\}$ with a scenario tree $\{x_{t,n}, n \in N_t, 0 \leq t \leq T\}$, we can write down the above LP model in the discrete form as below,

$$v^*_{t,n} = \underset{\{v_{t,n}\}}{\text{Min}} \ v_{t,n} \tag{3.2.25}$$

$$s.t. \quad v_{t,n} \geq h(x_{t,n}), \tag{3.2.26}$$

$$v_{t,n} \geq \frac{1}{1+r\,dt} \sum_{m \in C(n)} p_{n,m} v^*_{t+1,m} \quad . \tag{3.2.27}$$

In above formulation, if all nodes of the scenario tree are uniquely ordered and indexed, we can ignore the time index in implementing the model. However, to avoid confusion, we still leave it here. The catch of this formulation is that the calculation of continuation values in the RHS of (3.2.27) depends on a sequence of option values one step later conditional on the current state. Each future option value $v^*_{t+1,m}$ recursively defines a new optimization of the same formulation (3.2.25)~(3.2.27) with proper time indices and states. Motivated by the observation from the numerical example in section 3.1.1, we can expand all dependent LP optimal stopping models and pack them together to write down the following LP to price the option value at any node (*t*, *n*) with $n \in N_t$ and $0 \le t \le T-1$,

$$v^*_{t,n} = \underset{\{v_{t,n}\}}{\text{Min}} \ v_{t,n} \tag{3.2.28}$$

$$s.t. \quad v_{t',n'} \ge h\left(x_{t',n'}\right), \quad v_{t',n'} \ge \frac{1}{1+r\,\Delta t}\sum\nolimits_{m\in C(n')} p_{n',m} v_{t'+1,m}\,, \tag{3.2.29}$$

$$\forall n' \in N_{t'}\left(n\right),\ t' = t,\cdots,T-1,$$

$$v_{T,n'} \ge h\left(x_{T,n'}\right),\ \ v_{T,n'} \ge 0,\ \forall n' \in N_T\left(n\right), \tag{3.2.30}$$

where $N_{t'}\left(n\right)$ is a newly introduced node set and it only includes the nodes that are at stage *t'* and in the subtree rooted at node (*t, n*).

The model above gives us the option price calculated at any node of the scenario tree. The option price at the initial time *t=0* and root node *n=0* then is therefore given by

$$v^*_{0,0} = \underset{\{v_{t,n}\}}{\text{Min}} \ v_{0,0} \tag{3.2.31}$$

$$s.t. \quad v_{t,n} \ge h\left(x_{t,n}\right), \quad v_{t,n} \ge \frac{1}{1+r\,\Delta t}\sum\nolimits_{m\in C(n)} p_{n,m} v_{t+1,m}\,, \tag{3.2.32}$$

$$\forall n \in N_t\left(n\right),\ t = 0,\cdots,T-1,$$

$$v_{T,n} \ge h\left(x_{T,n}\right),\ \ v_{T,n} \ge 0\,, \quad \forall n \in N_T\,, \tag{3.2.33}$$

The following Proposition 3.1 shows that the LP model (3.2.31)~(3.2.33) does model the optimal stopping problem (3.2.21) and the optimal solution of the model gives the option price and an optimal exercising strategy.

Proposition 3.1 Along any path in the scenario tree, if the option has not been exercised prior to and at some node (*t, n*), any optimal solution v^* should guarantee that the inequality corresponding to the holding decision at each node between the root node and node (*t, n*) on this path mus be binding. So we have $v^*_{t,n} = \frac{1}{1+r\,\Delta t}\sum_{m\in C(n)} p_{n,m} v^*_{t+1,m}$ $= \text{Max}\left\{K - S_{t,n}, \frac{1}{1+r\,\Delta t}\sum_{m\in C(n)} p_{n,m} v^*_{t+1,m}\right\}$. The latter is essentially the discretization of the optimal stopping problem, $v^*_{t,n} = \text{Max}\left\{K - S_{t,n}, \frac{1}{1+r\,\Delta t}\tilde{\mathbb{E}}^{t,n}[v^*\left(t+\Delta t, m\right)]\right\}$.

[**Proof**] We can prove the proposition by contradiction. Without loss of generality, let's assume the option is never exercised at the beginning and is always exercised the first time that the intrinsic value is equal to or greater than the holding value. If the statement is not true for some optimal solution v^*, there exists such a node (*t, n*) in the scenario tree that the option is not exercised prior to the node and its holding inequality is not binding. Next we will show that such a solution is not optimal because we always can strictly improve it.

We can improve the given optimal solution by working backwards along the sub-path defined by node (*t, n*). From the statement and the above assumption, at the given optimal solution v^*, we have $v^*_{t,n} > K - S_{t,n}$ and $v^*_{t,n} > \frac{1}{1+r\,\Delta t}\sum_{m\in C(n)} p_{n,m} v^*_{t+1,m}$. We can reduce $v^*_{t,n}$ by a strictly positive number $\varepsilon_{t,n}$ without violating the feasibility. So

$\tilde{v}^*_{t,n} = v^*_{t,n} - \varepsilon_{t,n}$, $\varepsilon_{t,n} = v^*_{t,n} - \max\left\{K - S_{t,n}, \frac{1}{1+r\,\Delta t}\sum_{m\in C(n)} p_{n,m} v^*_{t+1,m}\right\} > 0$. The reduction leads to a strict reduction on the holding value of the parent node of (*t, n*) as well. Note $\sum_{m\in C(a(n))} p_{a(n),m} v^*_{t,m} = \sum_{m\in C(a(n))\setminus\{n\}} p_{a(n),m} v^*_{t,m} + p_{a(n),n} v^*_{t,n} > \sum_{m\in C(a(n))\setminus\{n\}} p_{a(n),m} v^*_{t,m} + p_{a(n),n} \tilde{v}^*_{t,n}$, where the left term is the undiscounted original holding value of node *a(n)*, the middle term is an expansion of the left term to isolate the node n, and the right corresponds to the updated holding value of node *a(n)* due to the update of the node *n*. For simplicity, we denote $\sum_{m\in C(a(n))} p_{a(n),m} \tilde{v}^*_{t,m} = \sum_{m\in C(a(n))\setminus\{n\}} p_{a(n),m} v^*_{t,m} + p_{a(n),n} \tilde{v}^*_{t,n}$ as the improved holding value. So $\varepsilon_{t-1,a(n)} = v^*_{t-1,a(n)} - \max\left\{K - S_{t-1,a(n)}, \frac{1}{1+r\,\Delta t}\sum_{m\in C(a(n))} p_{a(n),m} \tilde{v}^*_{t,m}\right\} > 0$. We can improve the parent of node *n*. We repeat this procedure while tracing back the path up to the root node and we can improve the option value by a strictly positive amount at the root node. So this contradicts to the optimality of the given solution v^*. Therefore the statement is true. ■

Lattice based models: the recombining scenario tree

The scenario tree we just defined is a general framework. In theory, we can use it to model all kinds of options, including the exotic path-dependent options. However, the generality of the scenario tree does come with a cost. The tree grows exponentially in the number of stages. Fortunately, some options' scenario trees show the recombining property and we can reduce the exponential growth rate to a linear growth rate. The resulting recombining trees are called lattices. The numerical example in §3.1.1 is an example of binomial lattice option pricing models. Since the explicit finite difference method is equivalent to the trinomial lattice model under proper conditions, we can create finite difference based LP option pricing models, in which the conditional expectations are calculated with finite difference methods.

3.2.3 Economic insights

We can interpret the LP option pricing model from the perspectives of both option sellers and buyers. Assume we have an American call (or put) and we want to sell it at a proper price. We need to determine how much we should charge the buyer. From the seller's view, the price must be sufficiently large to setup a self-financing portfolio to dynamically hedge the short position as against any exercising strategy of the option buyer. From the buyer's view, the option price is necessary to be competitive in the financial market. Otherwise, it is not worth holding the option.

The required minimal charge would be the fair option price. From this point of view, we can treat $v(t,\omega)$ as the total wealth of our hedging portfolio at any node (t,ω). Instead of explicitly constructing the replicating portfolio (we always can do this with a dynamic delta hedging), we only care about the total wealth at any node. This wealth either comes from immediate exercise or from continuation. Constraints (3.2.32) and (3.2.33) together impose the lower bounds on the value of the hedging portfolio. These lower bounds also express sufficient conditions for the hedging. The sufficiency means as long as the value of the hedging portfolio satisfies those bounds, the short position is almost surely covered (i.e., covered at each future state). The minimization of the objective of the problem says the price the writer is going to charge is a competitive (non-arbitrage) price, which rules out any arbitrage opportunity the writer herself could have. If the price is not competitive and the writer keeps some arbitrage opportunity, in efficient market, nobody would buy such a call. The supply-demand pressure will force the writer to set the price to be competitive and the writer would have no incentive to keep any arbitrage opportunity.

For any optimal solution to the LP option pricing model, the optimality ensures that at least one of the two inequalities at each node must be binding. We can prove this by contradiction. If there is a node such that the present option price is strictly greater than the continuation value and the intrinsic value, then we can decrease the present price by a strictly positive amount to make at least one inequality at this node binding. The reduction will strictly improve the old optimal solution – this is a contradiction.

Another advantage of the LP model is we can easily perform the sensitivity analysis. We can even do more based on the parameterized LP theory to identify a range for the considered parameters, within which the optimal exercising rule will not change.

3.2.4 An application: sequential investment problem

Long-term investment projects are usually carried out over several phases. Investors can incrementally invest such projects to distribute risks to different contingences over time. We can model such phased investment opportunities as sequential compound options, where some options' payoffs rely on the value of other options. The following example assesses a compound option with our LP models.

Assume the stock follows the same process as Figure 3-1. We have two options in sequence. The first is an American call with a strike price $K_1 = \$1$. It is immediately available and expires in two months. The exercise of the call buys the investor the second option, which is an American put on the underlying stock. The put is the same as Figure 3-1 except that it is activated at t = 2 only if the call is exercised. The compound option can be assessed with a lattice model as Figure 3-6 (a) and (b).

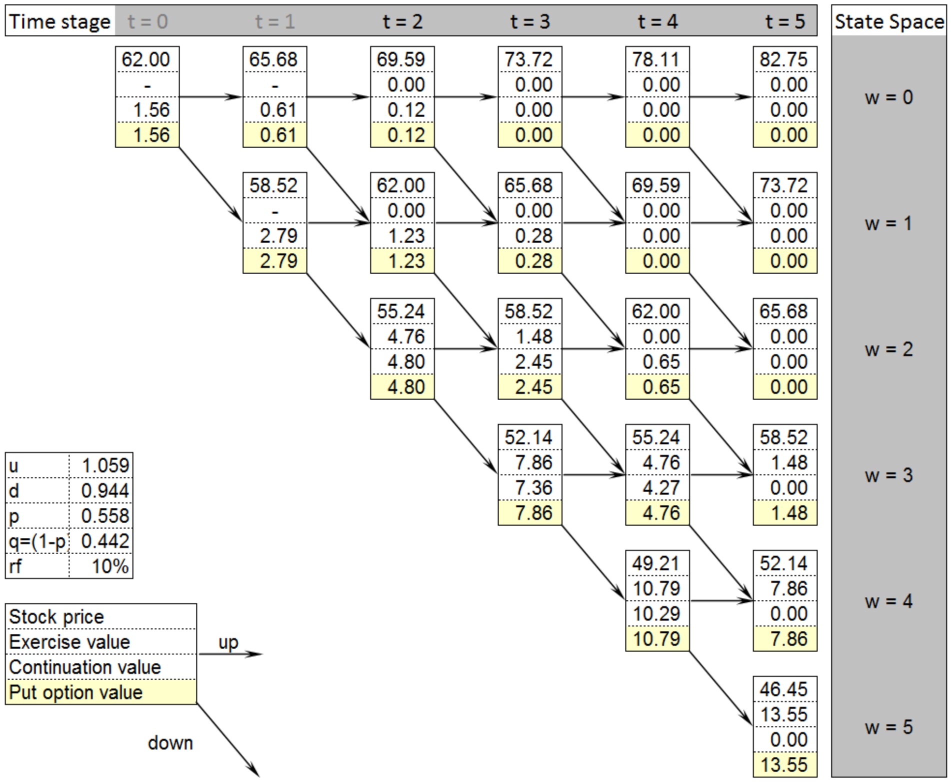


Figure 3-6 (a) The underlying American put (K_2 = \$60) is activated at t = 2 if the call (Figure 3-6(b)) is exercised by t = 2. Since it is not optimal to exercise the put prior to t = 3, the put are the same as Figure 3-1.

Since the call's payoff is contingent on the underlying put option value, we evaluate the put option first, which is shown in Figure 3-6 (a). Since the put is not activated until the end of the second month, the exercise values are all ignored and only continuation values matter prior to t = 2. However, since it is not optimal to exercise the put prior to t = 3, the put still has the same values as Figure 3-1. After we have calculated the put values, we can assess the call option, which is shown in Figure 3-6 (b).

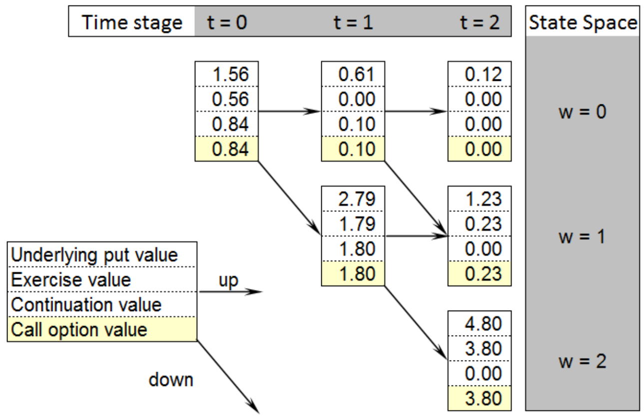


Figure 3-6 (b) The exercise of the American call (K_1 = \$1), which expires at t = 2, buys the investor the right to exercise the underlying American put.

So we have two choices to invest the 5-month American put on the stock. Either we can pay \$1.56 to directly buy the put in the market, or we can buy it in two steps to reduce risks. First, we pay \$0.84 to buy the call written on the put. If the underlying stock goes high and the underlying put becomes valueless, we can abandon the call at t = 2. As a result, we only lose the investment of \$0.84, instead of \$1.56 to buy the put directly. If the underlying stock goes low enough and the underlying put becomes valuable, we can exercise the call at a cost of $K_1 = \$1.00$ to earn the underlying put.

Let $v^*(t,x)$ denote the call value at time $t \in [0, T_1]$ with $T_1 = 2$ (months) and in state x, where x is the logged stock price at t, i.e., $x = \log S(t)$. Let $u^*(t,x)$ denote the underlying put value at time $t \in [0, T]$ with $T = 5$ (months) and in state x. We can write down the optimal stopping formulations for the compound options as follows,

$$v^*(t,x) = \max\left\{ u^*(t,x) - K_1, \frac{1}{1 + r\,dt} \tilde{\mathbb{E}}^{t,x}\left[v^*(t+dt, x+dx) \right] \right\}, \quad \forall t \in [0, T_1], \quad x \in \mathbb{R}, \tag{3.2.34}$$

where

$$u^*(t,x)=\begin{cases}\dfrac{1}{1+r\,dt}\tilde{\mathbb{E}}^{t,x}\left[u^*(t+dt,x+dx)\right], & t\in[0,T_1),x\in\mathbb{R}\\ \max\left\{K_2-S(t,x),\dfrac{1}{1+r\,dt}\tilde{\mathbb{E}}^{t,x}\left[u^*(t+dt,x+dx)\right]\right\}, & t\in[T_1,T],x\in\mathbb{R}\end{cases}. \tag{3.2.35}$$

Notice, the exercise payoff of the call in (3.2.34) relies on the value of the underlying put. The put value $u^*(t,x)$ during the first phase $[0,T_1)$ only comes from its continuation value since the exercise right is not enabled yet during the phase. Since the options are not path-dependent, we can use a lattice similar to the numerical example in §3.1.1. Let $v_{t,n}$ and $u_{t,n}$ to denote the values of $v^*(t,x)$ and $u^*(t,x)$, respectively, at node (t, n), where $S_{t,n}=e^{x_{t,n}}$. We can formulate the compound sequential option pricing problem as the following LP model,

$$v^*_{0,0}=\underset{\{v_{t,n},y_{t,n}\}}{\text{Min}}\ v_{0,0} \tag{3.2.36}$$

$$s.t.\quad v_{t,n}\ge u_{t,n}-K_1,\quad v_{t,n}\ge\frac{1}{1+r\,\Delta t}\left(p\cdot v_{t+1,n}+q\cdot v_{t+1,n+1}\right), \tag{3.2.37}$$

$$\forall n=0,\cdots,t,\ t=0,\cdots,T_1-1,$$

$$v_{T_1,n}\ge u_{T_1,n}-K_1,\ v_{T_1,n}\ge 0,\quad \forall n=0,\cdots,T_1,\ T_1=2, \tag{3.2.38}$$

$$u_{t,n}\ge\frac{1}{1+r\,\Delta t}\left(p\cdot u_{t+1,n}+q\cdot u_{t+1,n+1}\right), \tag{3.2.39}$$

$$\forall n=0,\cdots,t,\ t=0,\cdots,T_1-1,$$

$$u_{t,n}\ge K_2-S_{t,n},\quad u_{t,n}\ge\frac{1}{1+r\,\Delta t}\left(p\cdot u_{t+1,n}+q\cdot u_{t+1,n+1}\right), \tag{3.2.40}$$

$$\forall n=0,\cdots,t,\ t=T_1,\cdots,T-1,$$

$$u_{T,n}\ge K_2-S_{t,n},\ u_{T,n}\ge 0,\ \forall n=0,\cdots,T, \tag{3.2.41}$$

The model is convenient to implement and solve and just a slightly more complex than the standard American option pricing model. In this model, (3.2.37) and (3.2.38) model the immediate call option, whose payoffs are written on the underlying put which are modeled by (3.2.39) ~ (3.2.41). The optimal solution gives the option values to the compound options and corresponding exercising strategies.

3.3 Optimization Option pricing Models: Integer Programming

The LP models price financial options by seting up a series of lower bounds for the option value at each node and looking for the tightest one. All the binding bounds recursively work together to define the present option value. In this way, the binary exercise decisions are indirectly modeled. Sometimes, it may be preferable to directly work with the "go-no go" type decisions. In particular, sometimes, it is necessary to directly model the exercise decisions and associated exercise values, instead of use linear inequalities to bound option values. For example, the decision when to drill and how many wells to drill must be modeled by integer variables.

Wang and de Neufville (2004) develop an optimization model to evaluate lattice based real options with MIP techniques. In their models, the exercising and holding decisions are explicitly captured by binary variables. Brosch (2008) applies Wang and Neufville's ideas to model portfolios of real options. However, the original models of Wang and de Neufville (2004) are nonlinear due to the product form of decision variables to compute the continuation value at each node on the scenario tree. In order to solve large problems, additional constraints are required to decouple the product forms with a special MIP technique called "discrete alternative" and big-M constants. The linearization (or decoupling) unavoidably increases the model size and may deteriorate the numerical stability.

In this section, we develop a new integer programming (IP) based option pricing model. The model is a more general and powerful asset pricing framework. It may be used to price more general assets whose future cash flows are decision dependent. Our model is mainly motivated by Longstaff and Schwartz (2001) and Meier et al (2001) and can be readily applied to price more general options or other securities. To show its applications, we will model the compound sequential options already considered in §3.2.4 in the new approach, which turns out to be very simple and intuitive. Before we continue, let's give a brief review to the general idea behind our approach and use a small numerical example to show the idea.

3.3.1 A brief review to the general idea of asset pricing

The price of an asset equals its expected present value of future cash flows. In complete markets, the non-arbitrage argument leads to the risk neutral pricing method, in which the expectation is evaluated using risk neutral probabilities and the cash flows are discounted at the risk free rate. Particularly, for securities whose cash flows depend on the security owner's trading decisions, the owner must take optimal trading decisions to ensure that the resulting cash flows justify the asset price he has paid to own it.

For American options, the cash flows are the payoffs resulting from optimal exercising when the immediate exercise value dominates the continuation value. Given the optimal exercise strategy, we can estimate the option price in three steps. First, we generate a set Ω of sample paths for the underlying price process under the risk neutral measure. Second, we apply the optimal exercise strategies to each sample path $\omega \in \Omega$ to determine the optimal exercise epoch and the associated exercise payoff along the path. Finally, we average the discounted payoffs over all paths. The result is an unbiased estimate for the option price.

The sample paths are either simulated one by one as parallel paths as Longstaff and Schwartz (2001) or recursively generated and organized in a scenario tree structure as Meier et al (2001). The former approach applies a backward induction and approximates the continuation value by regressing it on polynomials in the underlying price at each time point *t*. The optimal exercise strategy hence can be numerically determined. The latter one only considers perpetual American calls whose optimal exercise strategies have closed form expressions in the underlying price.

Our approach is a generalization to Meier et al (2001) in that our model neither assumes the involved options to be perpetual nor requires any optimal exercise strategy beforehand. In essence, we explicitly model the exercise decisions and capture the exercise payoffs. Consequently, the optimal exercise strategy is generated as part of the solution of our succeeding IP based option pricing model.

3.3.2 Numerical example

Before presenting our IP based option pricing model, we first use a simple example to explain how to model explicit exercise decisions with binary variables and price the two-month American call in Section 3.2.4 with an optimization model. For simplicity, we ignore the pricing of the underlying put and assume the put values are given and denoted as $S_{t,n}$ at node (*t,n*) in Figure 3-7. Other parameters, such as exercise payoff $h_{t,n} = \max\{0, K_1 - S_{t,n}\}$ and risk neutral probabilities $p_{t,n}$, have been computed and attached to each node in the figure as follows.

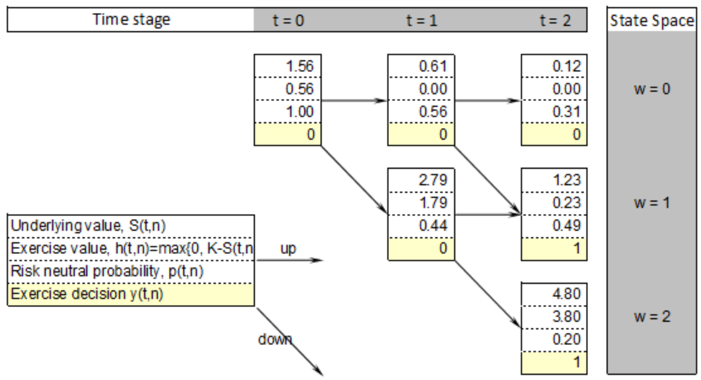


Figure 3-7 Price the American call with an integer programming model.

We define a binary decision variable $y_{t,n} \in \{0,1\}$ to model the exercise decision at node (t, n). Given an exercise strategy $\{y_{t,n},\ \forall t,n\}$, the expected present value of all resulting exercise payoffs is

$$
\begin{aligned}
&\sum\nolimits_{t=0}^{2}\left[\sum\nolimits_{n=0}^{t}\left(p_{t,n}\cdot e^{-r(t\cdot\Delta t)}\cdot h_{t,n}\cdot y_{t,n}\right)\right] \\
&= p_{0,0}\cdot h_{0,0}\cdot y_{0,0} \\
&+p_{1,0}\cdot e^{-r/12}\cdot h_{1,0}\cdot y_{1,0}+p_{1,1}\cdot e^{-r/12}\cdot h_{1,1}\cdot y_{1,1} \\
&+p_{2,0}\cdot e^{-r/6}\cdot h_{2,0}\cdot y_{2,0}+p_{2,1}\cdot e^{-r/6}\cdot h_{2,1}\cdot y_{2,1}+p_{2,2}\cdot e^{-r/6}\cdot h_{2,2}\cdot y_{2,2}\ .
\end{aligned}
\tag{3.3.1}
$$

Any feasible exercise strategy $\{y_{t,n},\ \forall t,n\}$ must satisfy one property that the option can be exercise at most once. As a consequence, along any sample path on the lattice in Figure 3-7, the option can be exercised at most once. Since there are four unique sample paths, we can express the feasibility conditions, one for each path, as

$$y_{00}+y_{10}+y_{20}\le 1, \tag{3.3.2a}$$

$$y_{00}+y_{10}+y_{21}\le 1, \tag{3.3.2b}$$

$$y_{00}+y_{11}+y_{21}\le 1, \tag{3.3.2c}$$

$$y_{00} + y_{11} + y_{22} \leq 1 . \tag{3.3.2d}$$

Since the uncertainties are captured by a lattice and the exercise decision variables are node-dependent, a delayed exercise decision can be made separately on different contingencies. In such a way, the constraints (3.3.2a~d) have implicitly restricted the exercise decisions to be non-anticipative. If we treat the binary exercise decisions as indicator functions of the exercise time (i.e., the decision variables in the optimal stopping problem (3.2.20)), the non-anticipativity property essentially models the stopping time property of the exercise time decisions.

So far, all ingredients required to model the option pricing problem are ready. The option price and the corresponding optimal exercise strategies can be found numerically by maximizing (3.3.1) subject to (3.3.2a~b) and $y_{t,n} \in \{1,0\}$, $\forall t, n$.

3.3.3 Mathematical development of the general pricing model

Let's reconsider the American option pricing problem (3.2.20). To model the exercise decision explicitly, we introduce a binary decision variable $y(\tau) \in \mathbb{B} \equiv [0,1]$ for each exercise time $\tau \in \mathcal{T}_{[t,T] \cup \{\infty\}}$. We define $y(\tau) = 1$ and $y(\tau) = 0$ to denote the "exercise" decision and "continuation" decision at τ, respectively. Since τ is a stopping time, the exercise decision variable $y(\tau)$ must be non-anticipative, which means the decision at τ, whether to exercise or not, must only depend on the observations of the underlying price process up to and include τ. In mathematical language, this requires the continuous time stochastic decision process $\{y(t), 0 \leq t \leq T\}$ to be adaptive to the filtration generated by the driving Brownian Motion process.

Additionally, we require that along any price sample path the option can be exercised at most once. Therefore, along any sample path, the integration $\int_t^T y(\tau)d\tau$ essentially counts the number of exercises along the path. It hence is well defined and equals either 1 or 0. Moreover, for any bounded function $h(\cdot): \mathbb{R} \to \mathbb{R}_+$ such that $|h(\cdot)| \le M$ for some big number M, integration $\left|\int_t^T e^{-r(\tau-t)} h(S(\tau)) y(\tau) d\tau\right| \le M$ is bounded and well defined along any sample path. So far the building blocks for our following model are all set.

To determine the option value $v^*(t,s)$ at time $t \in [0,T]$ and in state $S(t)=s$, we assume that the option has not been exercised up to that point. Otherwise, the option has a value of zero , i.e., $v^*(t,s)=0$. We can rewrite the pricing problem (3.2.20) as a new optimization problem as follows,

$$v^*(t,s) = \underset{\{y(\tau),\tau\in[t,T]\}}{\text{Max}} \tilde{\mathbb{E}}^{t,s}\left[\int_t^T e^{-r(\tau-t)} h(S(\tau)) \cdot y(\tau)\, d\tau\right] \tag{3.3.3}$$

$$\text{s.t.} \quad \int_t^T y(\tau) d\tau \le 1 \text{ almost surely,} \tag{3.3.4}$$

$$\{y(\tau)\} \text{ are binary and non-anticipative,} \tag{3.3.5}$$

where the stopping time restriction $\tau \in \mathcal{T}_{[t,T]\cup\{\infty\}}$ is substituted by constraint (3.3.5) and constraint (3.3.4) restricts that along any sample path the option can be exercised at most once. As a result, the integration inside the expectation of (3.3.3) contributes at most one discounted payoff to the objective value along any sample path. An exercise decision means giving up all future exercise opportunities and associated payoffs. In such a way, we have implicitly modeled the comparison and mutual exclusion between the immediate exercise value and the continuation value.

Although the non-anticipativity restriction in (3.3.5) seems abstract, it turns out to be easy to understand and natural to implement when the underlying price process is approximated by a scenario tree. In this case, non-anticipativity is equivalently to require the decisions to be adapted to the scenario tree. In other words, the decisions are node-dependent. With the introduction of a scenario tree to stochastic optimization models, the non-anticipativity is realized (Birge and Louveaux, 1997). In the following discussion, we assume that we have such a scenario tree as described in §3.2.2.

Since the asset price at any time t is determined by its future cash flows conditional on the present state, the option price $v^*_{t,m}$ at any stage-t node m, $m \in N_t$, in the scenario tree is determined by payoffs from the nodes in the sub-tree rooted at node m. We introduce a subset $N^{(m)}_\tau \subset N_\tau$ to denote the stage-τ nodes which reside on the sub-tree rooted at node m at stage t ($t \le \tau$). (Recall that all nodes in the scenario tree are contiguously numbered with unique integers. The node index contains the stage information.) So the family of node sets $\left\{N^{(m)}_\tau, t \le \tau \le T\right\}$ contains all nodes on the sub-tree rooted at node m. The subtree forms a sub-filtration conditional on the state represented by node m.

To express the constraints (3.3.4) in explicit form, we need a way to represent the sample paths in the scenario tree and any subtrees. We define a set Ω to enumerate all sample paths in the scenario tree. Each path $\omega \in \Omega$ consists of T+1 nodes ordered in the time index sequence. Therefore, the operator $\omega(t)$ returns the time-t node in the path ω. We introduce a subset $\Omega^{(m)}$ of Ω to address all the paths passing node m and composing the sub-tree rooted at node m. Therefore, we can write down the optimization model to price the option price at node m,

$$v^*_{t,m} = \underset{\{y_{\tau,n}\}}{\text{Max}} \sum_{\tau=t}^{T} \left[e^{-r(\tau-t)\Delta t} \sum_{n \in N^{(m)}_\tau} \left(p_{\tau,n} h_{\tau,n} y_{\tau,n} \right) \right] \tag{3.3.6}$$

$$s.t.\ \sum_{\tau=t}^{T} y_{\tau,\omega(\tau)} \leq 1,\ \ \forall\ \omega \in \Omega^{(m)}, \tag{3.3.7}$$

$$\left\{y_{\tau,n}\right\} \text{ are binary variables,} \tag{3.3.8}$$

where $h_{\tau,n} = h\left(S_{\tau,n}\right)$ calculates the immediate exercise payoff at the time-τ node n. Since the node indices are uniquely numbered, we can drop the time indices of the parameters and decision variables in the above model. However, to make the model clear and easy to understand, we keep the time indices.

The optimal value of the model (3.3.6) ~ (3.3.8) gives the option price at node m. Clearly, when $t = 0$ and $m = 0$, the IP model above determines the initial option price and associated optimal exercise strategy.

This formulation provides direct accesses to exercise decisions and thus presents a powerful and flexible approach to model more general option pricing problems. We will apply this approach to model and price the previous compound sequential options.

3.3.4 Pricing compound sequential options with the IP model

In §3.2.4, we present two approaches to price the compound sequential options. One is the traditional approach which uses a lattice-based backward induction. The other is an optimization method based on our linear programming option pricing model.

In this section, we take a more fundamental view to the same pricing problem in that we directly model exercise decisions and associated cash flows of the asset. The model provides an example to the general idea of asset pricing stated before.

First, we imitate the previous procedure to model the compound sequential option pricing problem as a continuous time optimization problem similar to (3.3.3) ~ (3.3.5).

Then, we approximate the continuous time pricing problem with a scenario tree based stochastic integer program.

Since two types of exercise decisions come in sequence, we introduce two binary decision variables $y(\tau), \tau \in [0, T_1]$ and $z(\tau), \tau \in [T_1, T]$ to denote the exercising of the call and the underlying put, respectively. We introduce two deterministic functions $h(S(\tau))$ and $g(S(\tau))$ to denote the two respective exercise payoffs. Since the exercise of the call simply buys the investor the underlying put and does not incur other cash flows, the cash flow associated to exercising the call is $h(x) \equiv -K_1, \forall x \geq 0$. Since exercising the underlying put will sell the stock at the put strike K_2, the payoff generated by exercising the put is $g(x) = (K_2 - x)^+, \forall x \geq 0$. So the price of the compound sequential options is determined by the optimal value of the following problem,

$$v^* = \underset{\substack{\{y(\tau),\tau\in[t,T_1]\},\\ \{z(\tau),\tau\in[T_1,T]\}}}{\text{Max}} \tilde{\mathbb{E}}\left[\int_0^{T_1} e^{-r\tau} h(S(\tau)) \cdot y(\tau)\, d\tau + \int_{T_1}^{T} e^{-r\tau} g(S(\tau)) \cdot z(\tau)\, d\tau \,\middle|\, S(0) = s\right] \tag{3.3.9}$$

$$\text{s.t.} \quad \int_0^{T_1} y(\tau) d\tau \leq 1 \text{ almost surely,} \tag{3.3.10}$$

$$\int_{T_1}^{T} z(\tau) d\tau \leq \int_0^{T_1} y(\tau) d\tau \text{ almost surely,} \tag{3.3.11}$$

$$\{y(\tau), z(\tau)\} \text{ are binary and non-anticipative.} \tag{3.3.12}$$

Similarly, constraint (3.3.10) ensures that the call can be exercised at most once. Constraint (3.3.11) models two relationships. First, the underlying put can be exercised only if the call has been exercised previously, under all situations. Second, the put can be exercised at most once given the call has been exercised. Again, constraint (3.3.12) requires that the binary exercise decisions only depend on revealed information.

If a scenario tree is available to approximate the underlying stock price process, we can approximate the above continuous time optimization problem with a discretized stochastic integer program as follows. We will use the same notations as those used in model (3.3.6) ~ (3.3.8). The model is,

$$v^* = \underset{\{y_{\tau,n}, z_{\tau,n}\}}{\text{Max}} \ \sum_{\tau=0}^{T_1}\left[e^{-r\tau\Delta t}\sum_{n\in N_\tau}\left(p_{\tau,n}h_{\tau,n}y_{\tau,n}\right)\right]+\sum_{\tau=T_1}^{T}\left[e^{-r\tau\Delta t}\sum_{n\in N_\tau}\left(p_{\tau,n}g_{\tau,n}z_{\tau,n}\right)\right] \tag{3.3.13}$$

$$\text{s.t.} \qquad \sum_{\tau=t}^{T} y_{\tau,\omega(\tau)} \le 1, \ \ \forall \ \omega \in \Omega_1, \tag{3.3.14}$$

$$\sum_{\tau=T_1}^{T} z_{\tau,\omega(\tau)} \le \sum_{\tau=0}^{T_1} y_{\tau,\omega(\tau)}, \ \ \forall \ \omega \in \Omega, \tag{3.3.15}$$

$$\left\{y_{\tau,n}, z_{\tau,n}\right\} \ \text{are binary,} \tag{3.3.16}$$

where $h_{t,n} = -K_1$, $g_{t,n} = \left(K_2 - S_{t,n}\right)^+$, Ω_1 denotes the set of sample paths of the subtree corresponding for the first phase, and Ω stands for the set of complete sample paths.

3.4 Portfolios of Real Options: A Dynamic Capital Budgeting Model

In this section, we develop a portfolio optimization model to manage multiple oil projects, which are modeled as real options. We use this model to study the dynamic capital budgeting problem, in which each project can be sequentially developed. We will use the model to find optimal strategies to prioritize and develop proven but undeveloped oil fields, subject to budget constraints and market uncertainty.

Each field is treated as an investment project to be developed in a two-phase procedure. In the first phase, some number of wells are drilled and start production. This initial production plan may be modified in the second phase to either accelerate or decelerate the depletion. The goal is to optimally allocate the budget to deplete the oil fields so that the total ENPV of the selected oil fields is maximized.

We will use the model to find the optimal drilling and production plans. First, we present a simple model to value completed oil projects. The model serves as a fundamental building block to calculate the option values to develop a field. Second, the field development is modeled as compound sequential options by our optimization framework. Finally, we model a number of oil projects and impose budget constraints. The budget constraints enforce risk sharing and synergy effects among projects. Consequently, the model captures project dynamics and interactions plus market risks.

3.4.1 Oil project valuation model

A simple way to evaluate an oil field is to use the exponential decline model to forecast production rates and assume no operational change thereafter. We ignore the economic limit of production and allow production to continue until the last drop of oil is extracted. More general models considering the economic limit and temporary shut-ins are available in Dixit and Pindyck (1994), though they are much more complicated.

We assume that the crude oil spot price follows the GBM process (3.1.1) with positive convenience yield (i.e., $\delta > 0$) and the unit production cost c is constant. Given time-t production rate q, decay rate λ, and current oil price s, the project value $V(s,q,\lambda)$ is given by the risk neutral pricing method as follows,

$$
\begin{aligned}
V(s,q,\lambda) &= \tilde{\mathbb{E}}^{t,s}\left[\int_t^\infty e^{-r(u-t)} q(u)\big(S(u)-c\big)du\right] \\
&= \int_t^\infty e^{-r(u-t)} q e^{-\lambda(u-t)} \tilde{\mathbb{E}}^{t,s}\left[S(u)\right]du - \int_t^\infty e^{-r(u-t)} q e^{-\lambda(u-t)} c\,du \\
&= \int_t^\infty e^{-r(u-t)} q\cdot e^{-\lambda(u-t)} s\cdot e^{(r-\delta)(u-t)}du - qc\int_t^\infty e^{-(r+\lambda)(u-t)}du \\
&= \frac{qs}{\delta+\lambda} - \frac{qc}{r+\lambda} = q\left(\frac{s}{\delta+\lambda} - \frac{c}{r+\lambda}\right),
\end{aligned}
\tag{3.4.1}
$$

where the time $u \geq t$ production rate $q(u) = q \cdot e^{-\lambda(u-t)}$. In the third equation above, we have used the relationship $\tilde{\mathbb{E}}^{t,s}\left[S(u)\right] = s \cdot e^{(r-\delta)(u-t)}$.

At any time *t*, once we know the well number *k* and the reservoir pressure $\overline{p}_t$, we can forecast the field production rate $q_t = q(\overline{p}_t, k)$ and the decline rate $\lambda_t = \lambda(k)$ with the tank model as follows,

$$J(k) = \frac{0.00708\, h\, k_0}{\mu_0\left(\frac{1}{2}\ln\frac{A}{r_w^2 C_A k} + 5.75 + s\right)} \quad \text{(rb/psi/day)}, \tag{3.4.2}$$

$$\lambda(k) = \frac{365\, kJ(k)}{V_p\, c_t} \quad \text{(1/year)}, \tag{3.4.3}$$

$$q(\overline{p}_t, k) = \frac{\lambda(k) V_p\, c_t}{B_0}\left[\overline{p}_t - \overline{p}_{wf}\right] \quad \text{(STB/year)}. \tag{3.4.4}$$

3.4.2 Two-phase dynamic oil field investment decision model

As a simple example, we consider a two-phase oil field depletion model. In the first phase, production facility will be constructed and some number of wells will be drilled and start production. The first phase investment buys the investor an option in the second phase to adjust the depletion plan. Contingent on the first phase field performance and latest oil price information, the second phase option allows the production to be either accelerated by drilling additional wells or decelerated and even abandoned by shutting in some or all wells. We can apply the sequential compound option pricing framework to evaluate the phased development of oil fields.

For simplicity, we make the following assumptions. First, we assume all geological uncertainties have been revealed and we only consider market uncertainty. Second, we ignore the construction time and assume that production starts immediately after wells are drilled. Third, in each phase, the drilling decision (or shut-in decision for the second phase) occurs just once but the decision is deferrable within the phase. Finally, we use the tank model to forecast annual production rates. Particularly, all wells with nonzero production rate after the last decision period will continue production at the exponentially declining rate forecasted by the tank model in perpetuity.

We model the development of each oil field as compound sequential options. The first-phase option is a two-year American call on the underlying project value with the capital expenditure as the strike price. The exercise decisions are when to install the processing facility and how many wells to be drilled. Contingent on the exercise of the first option, the second-phase option is to decide whether and when to drill additional wells or to shut in any production wells. The second option is also an American call on the residual value of the project. If the second option is exercised, the investor will receive the changed residual project value with the remaining project carried on under adjusted depletion plans, at the cost of giving up the original residual value. The strike price of the second option is a floating strike and equals the original residual project value. The project values and the residual project values can be estimated by the formula (3.4.1) ~ (3.4.4) with properly calculated reservoir pressure and production well number.

One interesting observation about the formula (3.4.1) is that under the foregoing assumptions, the project value for any oil field with known geological parameters doesn't depend on the time parameter but only depends on the present production rate, the decline rate, and the present oil spot price. If the first phase drilling decision is to drill k wells at time t in price state $S(t) = s_t$, the first phase option payoff is given by

$$h(s_t, q_t, \lambda_k) \equiv V(s_t, q_t, \lambda_k) - I_1, \tag{3.4.5}$$

where $S(t)=s_t$, $q_k=q(\overline{p}_0,k)$ and $\lambda_k=\lambda(k)$ as determined by (3.4.2) ~ (3.4.4). Here, since the field has not been depleted before, the reservoir pressure at the first drilling time *t* is still the initial reservoir pressure $\overline{p}_0$. After the drilling, the reservoir pressure will drop according to the tank model at the compound decline rate $\lambda_k=\lambda(k)$.

Now we turn to computing the incremented payoff of the second option at some time τ later than *t*, and $\Delta t=\tau-t>0$. The exercise of the second option changes the number of producing wells from *k* to k'. We denote $q_{k,k',\Delta t}$ as the field production rate after exercising the second option. The tank model hence forecasts $q_{k,k',\Delta t}$ with

$$q_{k,k',\Delta t}=\frac{\lambda_{k'}V_p c_t}{B_0}\left(\overline{p}_0-p_{wf}\right)e^{-\lambda_k\Delta t}=\frac{\lambda_{k'}}{\lambda_k}q_{k,k,\Delta t}, \tag{3.4.6}$$

where $q_{k,k,\Delta t}$ is the production rate immediately before exercising the second option and

$$q_{k,k,\Delta t}=\frac{\lambda_k V_p c_t}{B_0}\left(\overline{p}_0-p_{wf}\right)e^{-\lambda_k\Delta t}=q_k e^{-\lambda_k\Delta t}. \tag{3.4.7}$$

Formulae (3.4.6) and (3.4.7) forecast the production rates immediately after and before changing the initial production plan, respectively. Formula (3.4.7) essentially describes the simple static tank model since the well number is yet changed.

The incremental payoff $g(\Delta t,s_\tau,k,k')$ at τ due to exercising the second-phase option equals the difference between the adjusted residual project value and the original residual project value minus the adjustment capital expenditure I_2, so

$$g(\Delta t,s_\tau,k,k')=V\left(s_\tau,q_{k,k',\Delta t},\lambda_{k'}\right)-V\left(s_\tau,q_{k,k,\Delta t},\lambda_k\right)-I_2, \tag{3.4.8}$$

where $V\left(s_\tau, q_{k,k',\Delta t}, \lambda_{k'}\right)$ is the adjusted residual project value and $V\left(s_\tau, q_{k,k,\Delta t}, \lambda_k\right)$ the original residual project value.

So far we have all formulae (3.4.1) ~ (3.4.8) to determine the options payoffs. We can move on to setup the optimization model to assess oil projects.

Indices and index sets

$\mathbb{I}$: the set of oil projects, each for one oil field, superscripted by *i*.

$\mathbb{T}$: the set of decision stages, $\mathbb{T} = \{0, \cdots, T\}$ and *T*=5 (years), subscripted by *t*.

$\mathbb{T}_1$, $\mathbb{T}_2$: the sets of decision stages for both phases, $\mathbb{T} = \mathbb{T}_1 \cup \mathbb{T}_2$, $\mathbb{T}_1 = \{0, \cdots, T_1\}$, $T_1 = 2$, and $\mathbb{T}_2 = \{T_1 + 1, \cdots, T\}$. We use the convention that $t \in \mathbb{T}_1$ and $t' \in \mathbb{T}_2$.

$\mathbb{N}$: the node set of the binary scenario tree, $\mathbb{N} = \{0, \cdots, 2^{T+1} - 2\}$, subscripted by *n*. If $2^{t+1} - 2 \le n \le 2^t - 1$, the index *n* indicates a stage-*t* node. We use the convention that *n* denotes a first phase node and n' denotes a second phase node.

$\mathbb{N}_t$: the time-*t* subset of $\mathbb{N}$, $\mathbb{N}_t = \{2^t - 1, \cdots, 2^{t+1} - 2\}$ and $\mathbb{N} = \bigcup_{t \in \mathbb{T}} \mathbb{N}_t$.

$\mathbb{K}$: the full set of drilling/production plans, $\mathbb{K} = \{0, \cdots, K\}$, subscripted with *k*. It coordinates with the parameters $\{Nw_k^i\}$ to determine the number of production wells.

$\mathbb{K}_1$: a subset of $\mathbb{K}$. It only consists of the first-phase drilling plans. We use the convention that $k \in \mathbb{K}_1$ and $k' \in \mathbb{K}$.

There is a special property to quickly retrieve the (immediate) predecessor and successors of any node on an orderly symmetric binary tree. If the tree nodes are numbered by a breadth-first-search with the root numbered by 0, the two child nodes of any non-leaf node *n* are indexed by 2*n*+1 and 2*n*+2, respectively. Conversely, the immediate predecessor of any non-root node *n* is $\lfloor (n-1)/2 \rfloor$. With this property, we can efficiently traverse all intermediate nodes along the path back to the root node.

Decision variables

v^i : the project value corresponding to contingent investment plan $\left\{x_{tnk}^i, y_{tnk,t'n'k'}^i\right\}$

$$x_{tnk}^i := \begin{cases} 1, & \text{to choose production plan } k \text{ for field } i \text{ at stage-}t \text{ node } n \\ 0, & \text{otherwise} \end{cases}, \quad t \in \mathbb{T}_1$$

$$y_{tnk,t'n'k'}^i := \begin{cases} 1, & \text{to switch to plan } k' \text{ at stage-}t' \text{ node } n' \text{ for field } i \text{ from the old} \\ & \text{plan } k \text{ chosen at an earlier stage-}t \text{ node } n \quad \left(t \in \mathbb{T}_1, t' \in \mathbb{T}_2\right) \\ 0, & \text{otherwise} \end{cases}$$

Parameters

Nw_k^i : the number of wells for production plan $k \in \mathbb{K}$. For simplicity, $Nw_k^i = k$.

$p_{t,n}$: the risk neutral probability for stage-*t* node *n*.

c_{op}^i (\$/STB): unit production cost.

c_{dr}^i and c_{shut}^i (\$/well): drilling cost and shut-in cost per well, respectively.

I_1^i and I_2^i (\$): capital expenditure for the first phase and second phase production, respectively.

$I_{1,k}^i$ (\$): overall first phase capital expenditure with plan *k*, $I_{1,k}^i = Nw_k^i \cdot c_{dr}^i - I_1^i$.

$I_{2,k,k'}^i$ (\$): overall second phase capital expenditure with plan k', given the first phase produces with plan *k*, $I_{2,k,k'}^i = \left(Nw_{k'}^i - Nw_k^i\right)^+ \cdot c_{dr}^i + \left(Nw_k^i - Nw_{k'}^i\right)^+ \cdot c_{shut}^i + I_2^i$.

q_k^i (STB/year): the first phase initial production rate with plan *k*. It is determined by the tank model (2.2.11) and (2.2.12).

$V_{t,n,k}^i$ (\$): the project value at stage-*t* node *n* if plan *k* is chosen and the second phase investment is forbidden.

$V_{t,k,t',n',k'}^i$ (\$): the residual project value at stage-t' node n' when the production plan is changed to k' from the first phase plan *k* determined at some stage-*t* node.

$q^i_{k,k',\Delta t}$ (STB/year): the production rate after the second phase plan k' has been chosen, given the first phase plan *k* has been operated for Δt (years).

$h^i_{t,n,k}$ (\$): the first phase option payoff at stage-*t* node *n* if plan *k* is chosen.

$g^i_{t,k,t',n',k'}$ (\$): the second phase option payoff at stage-t' node n' when the plan is changed to k' from the first phase plan *k* which was chosen at some stage-*t* node.

All other geological parameters required by the tank model in §2.2. We assume all cost parameters are constants and will not change over time. So the deferral of investment favors capital expenditure saving due to the discounting effect. However, the deferral disfavors the delayed project value as well due to the loss of convenience yield. This is similar to the early exercise value of American call options when the underlying stock price process has a positive dividend rate.

The optimization model

Once we have computed the option payoffs $\left\{h^i_{t,n,k}, g^i_{t,n,k,t',n',k'}\right\}$, we can implement the project valuation problem with our optimization based compound options pricing model. The solutions of the problem propose the optimal contingent plans to develop the oil field. The objective is to maximize the ENPV of the project subject to technique constraints and stochastic market conditions.

$$\max_{\{x,y\}} \; v^i = \sum_{t\in\mathbb{T}_1, n\in\mathbb{N}_t, k\in\mathbb{K}_1} e^{-rt} p_{t,n} h^i_{t,n,k} x^i_{t,n,k} + \sum_{\substack{t\in\mathbb{T}_1, n\in\mathbb{N}_t, k\in\mathbb{K}_1 \\ t'\in\mathbb{T}_2, n'\in\mathbb{N}_{t'}, k'\in\mathbb{K}}} e^{-rt'} p_{t',n'} g^i_{t,k,t',n',k'} y^i_{t,n,k,t',n',k'} \tag{3.4.9}$$

$$\text{s.t.} \quad \sum_{t\in\mathbb{T}_1, k\in\mathbb{K}} x^i_{t,\omega(t),k} \le 1, \qquad \forall \omega\in\Omega_1 \tag{3.4.10}$$

$$\sum_{k'\in\mathbb{K}} y^i_{t,\omega(t),k,t',\omega(t'),k'} \le x^i_{t,\omega(t),k}, \qquad \forall t\in\mathbb{T}_1, t'\in\mathbb{T}_2, k\in\mathbb{K}_1, \omega\in\Omega \tag{3.4.11}$$

$$\sum_{\substack{t\in\mathbb{T}_1,k\in\mathbb{K}_1,\\ t'\in\mathbb{T}_2,k'\in\mathbb{K}}} y^i_{t,\omega(t),k,t',\omega(t'),k'} \le 1, \qquad \forall\omega\in\Omega \tag{3.4.12}$$

$$\left\{x^i_{t,n,k}, y^i_{t,n,k,t',n',k'}\right\} \text{ are binary variables.} \tag{3.4.13}$$

where $h^i_{t,n,k} = V^i_{t,n,k} - I^i_{1,k}$ and $g^i_{t,k,t',n',k'} = V^i_{t,k,t',n',k'} - V^i_{t,k,t',n',k} - I^i_{2,k,k'}$. Involved parameters can be computed by the following formulae,

$$V^i_{t,n,k} = q^i_k\left(\frac{S_{t,n}}{\delta+\lambda^i_k} - \frac{c^i_{op}}{r+\lambda^i_k}\right), \tag{3.4.14}$$

$$V^i_{t,k,t',n',k'} = q^i_{k,k',t'-t}\left(\frac{S_{t',n'}}{\delta+\lambda^i_{k'}} - \frac{c^i_{op}}{r+\lambda^i_{k'}}\right), \tag{3.4.15}$$

$$q^i_{k,k',t'-t} = \frac{\lambda^i_{k'}V^i_p c^i_t}{B^i_o}\left(\overline{p}^i_0 - p^i_{wf}\right)e^{-\lambda^i_k(t'-t)} = \frac{\lambda^i_{k'}}{\lambda^i_k}q^i_k e^{-\lambda^i_k(t'-t)}. \tag{3.4.16}$$

$$q^i_k = \frac{365Nw^i_k J^i_k}{B^i_o}\left(\overline{p}^i_0 - p^i_{wf}\right) = \frac{\lambda^i_k V^i_p c^i_t}{B^i_o}\left(\overline{p}^i_0 - p^i_{wf}\right). \tag{3.4.17}$$

The objective (3.4.9) expresses the fact that the project value equals the ENPV of its future cash flows as a function of the sequential investment decisions $\left\{x^i_{t,n,k}, y^i_{t,n,k,t',n',k'}\right\}$. Constraints (3.4.10) and (3.4.12) impose the restrictions that both the first phase investment decision and the second phase decision can be carried out no more than once along any sample path. The constraint (3.4.11) simply enforces the logic consistency between the first phase and second phase decisions.

Equalities (3.4.14) and (3.4.15) estimate the (scenario tree) node contingent project value at corresponding first phase and second phase decision points, respectively. They are based on (3.4.1). Particularly, (3.4.15) gives the residual value of the remaining part of the project which is determined by the remaining cash flows after the completion of the second phase investment.

Equality (3.4.17) computes the initial production rate under the first phase production plan *k*. Equality (3.4.16) computes the initial production rate under the second phase production plan k' and after the first phase production has lasted for a period of length $\Delta t = t' - t$ under plan *k*. Specially, if $k' = k$, we have $\lambda_{k'}^{i} = \lambda_{k}^{i}$ and the equality (3.4.16) is degenerated to the simple exponential decline model,

$$q_{k,k,t'-t}^{i} = q_{k}^{i} e^{-\lambda_{k}^{i}(t'-t)} . \tag{3.4.18}$$

3.4.3 Dynamic capital budgeting model

Oil projects are capital intensive projects and usually last for a long time period to recover the capital expenditure. Therefore, proper prioritization and investment strategies of promising projects are significant to the success of upstream business.

In the remainder of this section, we develop a dynamic capital budgeting model to select the best project mix and optimally allocate an initial limited budget over time to sequentially develop selected oil fields so that the total ENPV of the project portfolio is maximized. Once we have setup the pricing model (3.4.9)~(3.4.14) for each project, we can conveniently model the dynamic capital budgeting problem as follows.

Given a set $\mathbb{I}$ of oil projects, our objective is to maximize the total ENPV,

$$v^{*} = \underset{\{(x^{i}, y^{i}), \forall i \in \mathbb{I}\}}{\mathrm{Max}} \sum_{i \in I} v^{i} , \tag{3.4.19}$$

where each project's ENPV v^{i} are defined by equation (3.4.9). The feasible contingent investment strategies for each individual project are characterized by constraints (3.4.10)~(3.4.13).

To model the dynamic capital budget, we take an initial budget B and sequentially allocate it to selected projects to fund their development. The budget constraint should guarantee that for any realization of future oil prices and corresponding development strategies, the present value of the incurred capital expenditure can not exceed the initial budget B. The budget constraint is therefore enforced path-wise,

$$\sum_{t\in\mathbb{T}_1,k\in\mathbb{K}_1}\left[e^{-rt}\sum_{i\in\mathbb{I}}I_{1,k}^{i}x_{t,\omega(t),k}^{i}+\sum_{t'\in\mathbb{T}_2,k'\in\mathbb{K}}\left(e^{-rt'}\sum_{i\in\mathbb{I}}I_{2,k,k'}^{i}y_{t,\omega(t),k,t',\omega(t'),k'}^{i}\right)\right]\le B,$$

$$\forall\omega\in\Omega. \qquad (3.4.20)$$

The budget constraint (3.4.20) captures the interactions and budget sharing among projects. The left hand side simply sums all discounted capital expenditure incurred by investment decisions of all projects over time. In fact, the constraint implicitly assumes that the any unused budget at any stage will accrue interests at the risk-free rate. In this sense, the constraint is essentially a compact form of the more general inventory-style dynamic rebalancing constraint, which has been widely considered in asset liability and management (ALM) literature such as Consigli and Dempster (1998).

The budget constraint (3.4.20) brings two challenging computational difficulties simultaneously into our model. The first difficulty is that the budget constraint is path-dependent since the budget availability in any state relies on the past capital expenditure and hence carries historical path information. As a result, the simple recombining tree is insufficient to model the budget constraint.

The second difficulty is that the budget constraint belongs to the notorious NP-hard knapsack constraint class. What is even worse is that we have one such constraint for each sample path. Although our interests in this dissertation is mainly focused on modeling project assessment and selection instead of designing efficient algorithms, we will propose some effective and easy-to-implement schemes to accelerate the solution procedure in the following numerical experiments and discussion part.

So far, we have defined the basic dynamic capital budgeting model for sequentially developed oil projects. More realistic restrictions can be added to this fundamental model to ensure more practical and meaningful results. We also can replace the objective function with one taking care of more appropriate risk measures.

3.4.4 Solution procedure

Large instances of the dynamic capital budgeting problem are difficult to solve for two reasons. The first reason comes from the knapsack-type budget constrains and the resulting non-recombining scenario tree, which causes the model size to grow exponentially in time stages. The second reason is due to the linearization of the dynamic tank model, in which production curves are (implicitly) forecasted as nonlinear functions of drilling decisions as explained in Section 2.2. The linearization by discretization unavoidably leads to the sharp increase in the number of discrete decision variables. Other linearization approaches like piecewise linear approximation also require a number of binary variables of similar magnitude since any linearization must be carried out along each sample path of the scenario tree. Both causes combine and pose challenging computational burdens to the optimization solvers.

We propose three simple methods to accelerate the solution procedure by tightening the IP model formulation, identifying nice solution structures for the solvers, and reducing the model size, respectively.

First, we can strengthen the original formulation (3.4.9) ~ (3.4.13) by substituting the forcing constraint (3.4.11) with an equivalent but tighter aggregated forcing constraint (3.4.21) as below,

$$\sum_{t' \in \mathbb{T}_2} \sum_{k' \in \mathbb{K}} y^i_{t,\omega(t),k,t',\omega(t'),k'} \leq x^i_{t,\omega(t),k}, \quad \forall t \in \mathbb{T}_1, k \in \mathbb{K}_1, \omega \in \Omega. \tag{3.4.21}$$

The left hand side of (3.4.21) aggregates all second phase drilling decisions which reside on the same sample path and are contingent on the same first phase drilling decision. In contrast, the left hand side of (3.4.11) only counts the second phase decisions at a specific time point, instead of at all second-phase time points along the path like (3.4.21). If we sum both sides of constraint (3.4.21) respectively over $t \in \mathbb{T}_1$ and $k \in \mathbb{K}_1$ for each sample path $\omega \in \Omega$, we will come up with the constraint (3.4.12), using (3.4.10). So with constraint (3.4.21), the constraint (3.4.12) becomes redundant. Therefore, in the strong formulation with constraint (3.4.21) present, we can drop both constraints (3.4.11) and (3.4.12). The new formulation is particularly efficient when we need solver large instances of the dynamic capital budgeting problem.

Second, the drilling decisions in both phases involve the choice of a drilling plan from pre-specified plan set $\mathbb{K}$ or $\mathbb{K}_1$. Since those plans are mutually exclusive to each other, a natural thought is to specify the binary decision variables $\left\{x^i_{t,n,k}, y^i_{t,n,k,t',n',k'}\right\}$ to be the so-called specially-ordered-set of type one (SOS1) variables over the index on set $\mathbb{K}$ or $\mathbb{K}_1$. Most commercial mixed integer programming solver can take advantage of this information to expedite the branch-and-bound procedure. Since the plan sets $\mathbb{K}$ and $\mathbb{K}_1$ contain mutually exclusive elements, we can append to them a null plan, which corresponds to drilling zero wells, to model the decision to defer the drilling decisions, or abandon the project (in the first phase) or give up the chance to adjust the production plan (in the second phase).

The last technique is to scan the pre-calculated payoff parameters $h^i_{t,n,k}$ and $g^i_{t,k,t',n',k'}$. If any of them has a non-positive value, we can fix the associated decision variables $x^i_{t,n,k}$ or $y^i_{t,n,k,t',n',k'}$ to zero. We would not exercise in those states anyway.

Our dynamic capital budgeting model is a multistage stochastic program. Like any other multistage stochastic programs, the single biggest numerical challenge comes from the exponential growth of the scenario tree. If we can control the growth rate of the scenario tree, we can effectively solve larger instances with more time stages. Fortunately, there is a general purpose scenario (tree) reduction code, called GAMS/SCENRED, provided in GAMS as a standard module. Since we will describe this scenario reduction code later in this dissertation, we will not discuss the details now.

With those techniques mentioned above, we can readily solve instances of moderate sizes. Since our interests in this dissertation are concentrated on modeling the problems of assessing and prioritizing option values embedded in oil projects, we will not delve into the design of more efficient special-purpose algorithms to solve the dynamic capital budgeting problem. We will leave this as a topic for future research.

3.4.5 Numerical experiments and economic insights

We implement the model in GAMS and solve it by calling the CPLEX solver. The computer on which we perform the numerical tests is a Lenovo T400 laptop equipped with an Intel® Core™2 Duo CPU P8600 @ 2.40GHz. The OS platform is Microsoft Windows 7.

Using the foregoing solution techniques, we can solve instances up to ten time periods and ten oil projects within three minutes on average with the relative termination tolerance (optcr) to be 5%. Those instances are reported to contain 0.8 million binary variables before pre-solve and a half million of binary variables after pre-solve. We can readily solve smaller instances with fewer time stages in just seconds or tens of seconds.

We can estimate the model size using big-O analysis. For each project $i \in \mathbb{I}$, the number of the first phase decisions $\left\{x_{t,n,k}^{i}, \forall n \in \mathbb{N}_1, k \in \mathbb{K}_1\right\}$ is $\sim \left|\mathbb{N}_1\right| \cdot \left|\mathbb{K}_1\right| \sim \left(2^{\left|\mathbb{T}_1\right|+1} - 1\right) \cdot \left|\mathbb{K}_1\right| \sim 2^{\left|\mathbb{T}_1\right|+1} \cdot \left|\mathbb{K}_1\right|$. Contingent on each first phase decision $x_{t,n,k}^{i}$, we have $\left|\mathbb{N}_2\right| \cdot \left|\mathbb{K}\right|$ second phase decisions $\left\{y_{t,n,k,t',n',k'}^{i}, \forall n' \in \mathbb{N}_2, k' \in \mathbb{K}\right\}$. Since we have $\left|\mathbb{I}\right|$ projects, $\left|\mathbb{T}\right| = \left|\mathbb{T}_1\right| + \left|\mathbb{T}_2\right|$, and $\left|\mathbb{N}_2\right| = \left|\mathbb{N}\right| - \left|\mathbb{N}_1\right| \le \left|\mathbb{N}\right|$, so a loose upper bound of the total number of the second phase decisions is

$$\left|\mathbb{I}\right| \cdot \left|\mathbb{N}_1\right| \cdot \left|\mathbb{K}_1\right| \cdot \left|\mathbb{N}\right| \cdot \left|\mathbb{K}\right| \sim \left|\mathbb{I}\right| \cdot 2^{\left|\mathbb{T}_1\right|+1} \cdot \left|\mathbb{K}_1\right| \cdot 2^{\left|\mathbb{T}\right|+1} \cdot \left|\mathbb{K}\right| \sim \left|\mathbb{I}\right| \cdot 2^{2\left|\mathbb{T}_1\right|+\left|\mathbb{T}_2\right|+2} \left|\mathbb{K}_1\right| \cdot \left|\mathbb{K}\right|,$$

which dominates the first phase decision. So the magnitude of the decision variables is up to $\left|\mathbb{I}\right| \cdot 2^{2\left|\mathbb{T}_1\right|+\left|\mathbb{T}_2\right|+2} \left|\mathbb{K}\right|^2$. For our instances with $\left|\mathbb{I}\right| = 10$, $\left|\mathbb{T}_1\right| = 3$, $\left|\mathbb{T}_2\right| = 7$, and $\left|\mathbb{K}\right| = 4$, the big-O analysis gives an upper bound of 2.6 millions to the problem size. We can see that it is the exponential growth of the scenario tree, measured by $2^{2\left|\mathbb{T}_1\right|+\left|\mathbb{T}_2\right|+2}$, that poses the main computational burden. A similar big-O analysis provides an estimated upper bound to the number of constraints.

From our numerical experiments, we found that once the investment time length is greater than some critical level, e.g. four years for our cases below, the objective value converges to some level and tends to be invariant to the increase of time periods. This is reasonable due to the conflicting dual effects of deferring investment decisions. From one perspective, deferring the investment not only postpones the capital expenditures, which are assumed to be constant, but also allows the decisions to be made more informatively. These factors favor deferring. However, from the other perspective, deferring the investment means losing the convenience yields during waiting, which also explains why the early exercise of American calls on a dividend-paying asset may be valuable. As a result, we can use small instances to continue our following discussions without losing accuracy.

The following numerical experiments are performed for an instance of four years with five projects, with each phase lasting for two years. We divide the four years into four time periods, each for one year. We want to find the best project mix and associated optimal contingent drilling plans to develop them. These oil fields are hypothetical and their parameters are randomly sampled from the base case (project i_1, studied in Section 2.4) and listed in Table 3-1 below.

Table 3-1: Geological and economic parameters for the five oil fields.

Parameters \ Oil fields			i1	i2	i3	i4	i5
A	Drainage area	acres	2084.000	1930.450	4589.399	3429.666	2442.786
rw	Well radius	ft	0.500	0.500	0.500	0.500	0.500
h	Reservoir thickness	ft	116.000	134.004	118.981	146.703	258.322
P0	Reservoir pressure	psi	3991.000	2903.517	6187.647	9963.226	6783.077
Pwf	Bottomhole pressure	psi	400.000	993.261	819.310	339.326	726.186
Bo	Oil formation volume factor	rb/STB	1.210	1.093	1.301	2.264	1.727
mu0	Oil viscosity	cp	10.000	12.834	12.677	8.498	8.852
ko	Permeability	mD	373.000	641.305	812.654	387.379	695.606
phi	Porosity	%	0.260	0.656	0.352	0.227	0.480
So	Oil saturation	%	0.720	0.455	0.875	0.516	0.529
Sw	Water saturation	%	0.280	0.545	0.125	0.484	0.471
sf	Skin factor	-	5.000	5.000	5.000	5.000	5.000
Ca	Shape factor	-	31.620	31.620	31.620	31.620	31.620
co	Oil Compressibility	1/Mpsi	50.000	86.426	98.658	89.684	74.061
cw	Water Compressibility	1/Mpsi	1.000	1.385	0.824	1.197	0.688
cf	Formation Compressibility	1/Mpsi	0.050	0.062	0.047	0.091	0.083
ct	Total compressibility	1/Mpsi	36.330	40.099	86.452	46.984	39.559
Vp	Pore volume	MMrb	487.604	1315.790	1489.248	886.654	2347.502
VOOIP	Original Oil In Place	MMSTB	290.144	547.279	1001.375	202.257	718.613
capex1	Initial capital cost	$ millions	250.000	471.557	862.824	174.273	619.185
capex2	Expansion capital cost	$ millions	150.000	282.934	517.695	104.564	371.511

Besides the capital expenditures listed in Table 3-1, we assume the following costs in Table 3-2 common for all five fields. To avoid detailed accounting issues, we ignore royalties and taxes and roughly include them in the operational cost.

Table 3-2 Operational and other costs

Parameters	Unit	Amount
Operational cost (OPEX)	$ / STB	45.00
Drilling cost	$ million / well	4.00
Premature shut-in cost	$ million / well	2.00

We can solve the resulting four-period model within just a few seconds on average. Next, we use the model to study an interesting problem of how budget levels impact the optimal drilling plans and the project prioritization. We keep other parameters fixed, increment the budget level from low to high, and solve the model under each budget level. The following Figure 3-8 shows the relationship between the optimal portfolio ENPV and the initial budget level.

The most apparent character of the curve is that it is made up of a sequence of progressively raised plateaus as the budget is increased. This discontinuity discloses one major difference between the portfolios of real projects and financial assets. Real projects, particularly for capital intensive ones like oil projects, are developed by a series of “go-no go” type investment decisions. Each decision usually consumes a lump sum of capital. On each plateau, the left-most point corresponds to the (most) efficient capital usage point among all investment opportunity on the plateau. Other points on the plateau occupy additional capitals but still return the same ENPV. Therefore, a slight increase in budget to the efficient point would not help to improve the overall

ENPV unless the increased budget is sufficient to drive the ENPV-budget pair jump to higher plateaus.

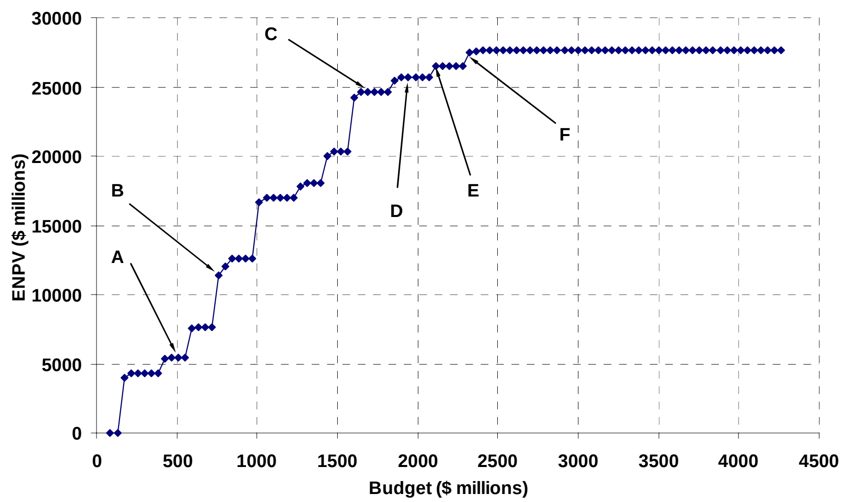


Figure 3-8 Portfolio ENPV versus budget curve. Each data point corresponds to a contingent plan to develop selected projects.

Figure 3-8 not only shows the relationship between portfolio ENPVs and budget levels in a wide range, but also identifies the critical budget level to reach a particular portfolio ENPV in an efficient way. In practice, we are more interested in questions like, for a given budget level, which projects should be selected, what are the optimal contingent drilling plans, and how the project mix and drilling plans evolve as the budget level varies. To answer those questions, we draw the following two figures, Figure 3-9 and Figure 3-10, and pick six different budget levels (A ~ F indicated on the curve in Figure 3-8) to explain how to use our dynamic capital budget model to find answers to the above questions. For simplicity, we call those points as budget-A or portfolio-A. These names are self-explained and should be clear.

Figure 3-9 details the portfolio composition of oil projects for each budget level. The figure draws individual project ENPVs of the optimal portfolio as functions of budget. We can see that the ENPV curve of each project is basically a horizontal line with slight fluctuations and sporadic broken segments. The line level represents the maximal ENPV that the project can reach, with or without budget restrictions. Any fluctuation falls below the line and represents that the project is operated under "non-optimal" drilling plans due to stringent budget constraints. Any broken segment indicates that the project is not included in the portfolio under the corresponding budget.

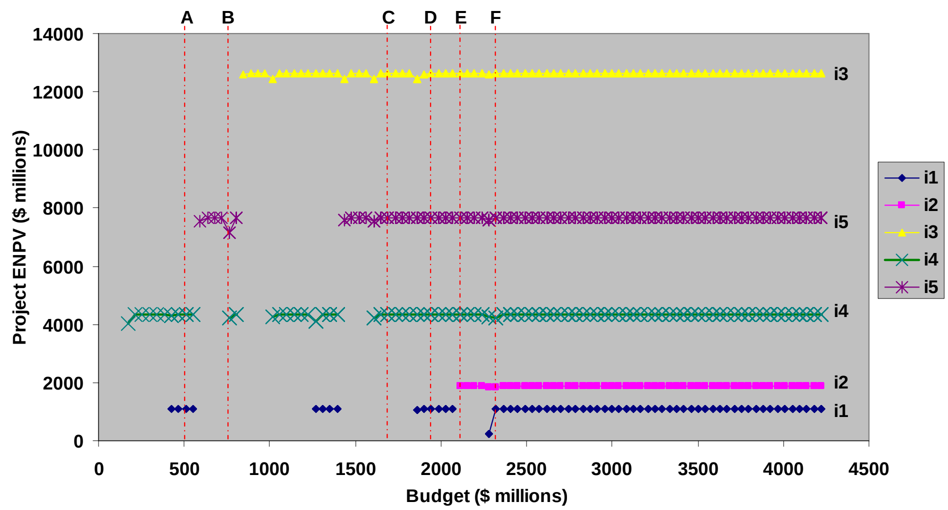
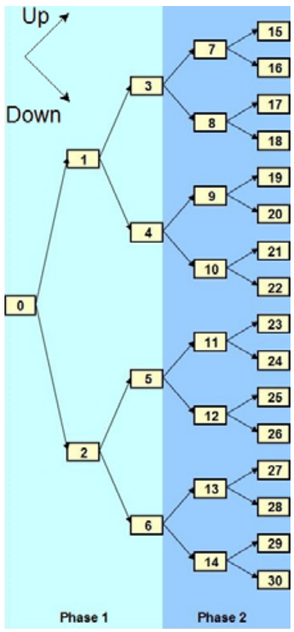


Figure 3-9 Individual project ENPV versus budget curve. This curve shows the project prioritization as the budget is incremented. The discontinuous parts of each project's value line imply that the project is dominated by other competitive project(s) mix under corresponding budgets.

In Figure 3-9, we draw a series of vertical dash lines (in red color) to indicate the positions of the six budget levels A ~ F. We call these lines as budget level lines or simply budget lines. If a budget line intersects with any project's ENPV line, that project will be included in the portfolio. For example, the budget line A intersects with the ENPV lines of project 1 and project 4. So under the budget level A, these two projects shall be included. As we move the budget line horizontally from left to right, the budget will rise and the budget line will sequentially intersect with different ENPV lines. From the intersection points, we can conveniently identify the project mix under each budget level and see how the project portfolio evolves as the budget changes. The detailed project mixes and associated optimal drilling plans for the six budget levels are summarized in Figure 3-10 as below.



Budget	Node	1	2	3	7	10	11	15	19	22	23	25
A	i1	1	1		5						3	
	i4	2	1		8		3					
B	i4	2	1				2			1		
	i5	1	1									2
C	i3	5	3	10							10	
	i4	2	1		8		3					
	i5	2	1					10			4	
D	i1	1	1		5						3	
	i3	5	3	10							10	
	i4	2	1		8		3					
	i5	2	1					10			4	
E	i2	3	1		6		4					
	i3	5	3	10							10	
	i4	2	1		6		3					
	i5	2	1					7			4	
F	i1	2	1			1					3	
	i2	3	1			2	4					
	i3	5	3	10							10	
	i4	2	1				3		5			
	i5	2	1		4						3	

(a) Binary scenario tree (b) Project mixes and optimal drilling plans

Figure 3-10 Project mixes and optimal contingent drilling plans for the sixe budget levels. (a) gives the two-phase four-period scenario tree structure; (b) shows the project mix and node-based drilling plans for each budget.

Figure 3-10 (a) shows the scenario tree for our example. The tree models the price evolution over the next four time periods. Each period represents one year. The four periods are divided into two investment phases, with one phase lasting for two periods. The nodes of the tree are orderly numbered as shown. Figure 3-10 (b) shows the optimal drilling plans by specifying how many wells to be drilled in a specific oil field in a scenario tree node.

For example, under budget level A, only field-1 and field-4 will be selected. The optimal drilling plans for them can be determined according to Figure 3-10 (b) as follows. At the beginning (i.e., node-0), no investment action is taken and any drilling decision should be held until the oil price becomes more favorable at a later time. After one year's waiting, the oil price may rise or drop. If the oil price rises and reaches node-1, we should drill one well in field-1 and two wells in field-4; otherwise, if the oil price drops to the level of node-2, we would drill only one well in each field. According to the drilling plans in Figure 3-10 (b), no other operational changes will be made in the remaining time of the first phase.

In the second phase, additional wells may be drilled and the capacity may be expanded if the residual initial capacity is insufficient. In our example, we assume the initial capacity is to drill totally ten wells. If more wells are to be drilled in the second phase, we should expand the capacity at the expansion costs. At the end of the period-3, if the oil price state is node-7, five additional wells will be drilled in field-1 and eight new wells will be drilled in field-2; if the oil price state is node-11. So no expansion is necessary for both fields. At the final investment stage, the end of the fourth period, only three wells will be drilled in field 1 if the market state is node-23.

Further analyses to the previous tables and graphs reveal further interesting properties of the optimal drilling plans. One interesting thing that we have just mentioned is that our model allows some projects to be developed in non-optimal drilling plans in the portfolio due to insufficient budgets. This is because our overall goal is to maximize the portfolio value, instead of optimizing the performance of each individual project. Therefore, this provides an example that we give up local optimality in order to achieve gains in the overall objective. The portfolio-B, the portfolio under budget-B, in Figure 3-9 is a good example. From the intersections on the budget line B, we can see that the portfolio-B consists of two projects, project-4 and project-5, both of which are operated sub-optimally since their corresponding ENPVs fall below their normal levels. However, if we look Figure 3-8, the portfolio-B is on a sharp rising edge of the ENPV-budget curve. A slight decrease in budget will significantly deteriorate the portfolio performance. Any attempt to improve the performance over portfolio-B requires additional budget. As a consequence, although those two projects are not developed optimally with respect to their respective potential, they together provide the most efficient use of the available budget.

This observation also verifies that our IP based option pricing model and dynamic capital budgeting model is an improvement and generalization over Meier et al (2001). In their model, each individual project must be optimally operated no matter what budget is available since their model explicitly relies on the optimal exercise rules for each project. Hence, Meier's model rules out all investment opportunity in the "gray" zone which actually may be feasible and beneficial to sacrifice partial local interests in the return of bigger global gains under stringent budgets.

Finally, we conclude this section with a comment regarding Figure 3-9. Although there are fluctuations and broken parts in the project ENPV lines, those events occur sporadically and only in narrow budget windows. The majority of these ENPV lines achieves the highest level. So once an oil field is included the portfolio, it is very much likely that it should be operated in the optimal state.

3.5 CONCLUSION

In this chapter, we have developed optimization based option pricing models and modeled portfolios of real options to deal with the dynamic capital budgeting problem. We then apply the real options based dynamic capital budgeting model to screen and develop explored oil fields over multiple periods and under the oil price uncertainty. We assume each oil field can be sequentially developed in two phases, an initial production phase and a succeeding production adjustment phase. Our objective is to find out the best oil project mix and the optimal drilling plans over time for a given initial budget while taking into account the market opportunity. We model and assess individual oil projects as compound sequential options with the exercise decisions being how many wells to be drilled and the resulting developed field being the underlying asset. To achieve this goal, we have developed optimization based option pricing models to model and solve the valuation problem for compound sequential options. Our option pricing models are general purpose frameworks and have many advantages over traditional approaches.

Chapter 4 Multistage E&P Portfolio Optimization Model

In Chapter 3, we considered the problems of evaluating and prioritizing multiple explored oil fields under budget constraints and market uncertainty. Assuming a complete market, we developed optimization based option pricing methods and real options based dynamic capital budgeting framework to deal with these problems. In this chapter, we move on to consider similar but more challenging and practical problems for unexplored oil fields, known as exploration and production (E&P) projects.

Unlike the explored oil fields, there is no complete market for E&P projects. Little information about the geological properties of an unexplored oil field, such as the chance to find oil, the reservoir size, and the production rates, is available at the early stages when to make acquisition and preliminary exploration decisions. Moreover, it is not uncommon that millions of dollars may have been spent on an oil field before it is proven commercially unfavorable due to inadequate reserves or low extractability caused by, e.g., low reservoir pressure. Consequently, E&P projects are among the most risky ventures in the oil industry and the geologic risk dominates the market risk. Traditionally, the best we can do is to attach a probability distribution to each random parameter, then perform various decision analyses, field by field (Newendorp, 2000), and finally allocate budget to them one after the other according to certain criteria of priorites. Recently, more and more people have realized the advantages and importance of assessing projects and managing their risk in a global portfolio perspective. In this chapter, we apply optimization techniques to model multistage E&P project portfolios.

Due to the disparate nature of uncertainty sources between proven and unexplored oil fields, we need a new way to evaluate and model unexplored oil fields, which must take into account various geological risks and the investor's attitude toward them. The resulting investment strategies should trade off the total rewards against the overall risk in some optimal way. We will pick a proper risk measure and use a mean-risk utility function to make sure that the trade-off is aligned to the investor's risk propensity.

It has been a prevailing practice in the oil industry to select and hold E&P projects of different risk profiles from diverse locations, in hope of effectively hedging both geological and geopolitical risks in a consistent way. Motivated by this observation and the recent work by Ball and Savage (1999) and Gustafsson and Salo (2005), we develop a multistage stochastic optimization model to manage a portfolio of E&P projects such that the portfolio return is optimally traded off against the expected downside risk.

The multistage stochastic optimization model requires a scenario tree to approximate the investor's belief about how the geological parameters of all E&P projects evolve over time. In this chapter, we will take a scenario tree as given and concentrate on formulating the multistage E&P portfolio optimization (MEPPO) model. The model will be a significant extension to the static EPPO model of Ball and Savage (1999). In this model, the investment decision process of each project is modeled as a decision tree with mixed integer programming (MIP) technique. In the next chapter, we will discuss how to generate the scenario tree. We pay special attention to how to model project dependences and incorporate statistical learning to refine the scenarios based on preliminary information and experts' assessments.

The organization of this chapter is arranged as follows. In §4.1, we provide a brief introduction to the valuation and organization of E&P projects. In §4.2, we model a typical E&P project as a decision tree and its uncertainties as a scenario tree. Since the model is designed to cover the full lifecycle of the E&P project, we will focus on the most critical decisions and uncertainties. In §4.3, we present the multistage E&P project portfolio optimization (MEPPO) model in details. In §4.4, we develop efficient methods to solve the MEPPO problem. Based on extensive numerical experiments, we discuss the model properties and identify major computational challenges for future extensions. Since the MEPPO model involves complicated decisions and contingencies, we apply different methods to assure its correctness and effectiveness. Finally, we propose a statistical approach to interpret the solutions of the MEPPO model.

4.1 INTRODUCTION

We consider problems involving portfolios of oil field exploration and production (E&P) projects. Each project is associated with a specific reservoir, and involves decisions and uncertain outcomes in several time periods. The current decision is to determine which projects to initiate now given a prescribed initial budget. Later decisions include reservoir appraisal (including its geologic structure and size of recoverable reserves), how much crude oil processing capacity to install, how many production wells to drill, how much to expand the processing facilities, and the timing and extent of further production drilling.

We focus here on modeling only the geological uncertainties associated with each reservoir - oil prices are either assumed constant over the planning horizon at a predetermined value or forecasted by a fixed forward curve. This conforms to the current practice of many oil companies. In addition, all costs are taken as deterministic. Revenues and costs from all projects are viewed as being deposited and withdrawn from a single account. The initial value of this account is the starting exploration budget for all projects. The model computes values for both current and future decisions simultaneously, for all projects being considered. Thus, decisions on which projects to initiate are made with the assurance that future decisions are also being optimized.

These decisions depend on the (uncertain) outcomes of the exploration process up to the time the decision is made, so the result is not just a single sequence of decisions for each project, but a complete contingency plan.

The following two subsections give a brief review to E&P project valuation methods and the organization of E&P projects, respectively.

4.1.1 Valuation of E&P projects

We treat the development of an E&P project as a sequential investment problem under uncertainty. The optimal decision at any stage relies on both the current state and the expected future outcomes when all subsequent decisions are made optimally contingent on the current state and decision. This property motivates us to model the investment under uncertainty problem as a stochastic dynamic programming (SDP) problem. Using a scenario tree to approximate the underlying uncertainties, we can formulate the problem as a decision tree problem. With mixed integer programming (MIP) techniques, we can further formulate the decision tree into a multistage stochastic optimization problem. Once we can model a single E&P project, we can construct project portfolios by using a "bigger" portfolio state tree whose state space spans all individual projects' states and pooling all cash flows together in each state.

This approach is progressively developed by Heidenberger (1996), Ball and Savage (1999), and Gustafsson and Salo (2005). Heidenberger's model is essentially a multi-period capital budgeting model. He models multiple risky projects as parallel decision trees and then formulates the tree-based capital budgeting problem into a mixed integer linear programming (MILP). The resulting MILP is solved to find optimal contingent strategies to develop selected projects. The objective function is to maximize the overall benefit (ENPV). Ball and Savage provide a static E&P portfolio optimization (EPPO) model to consider the investment and risk management of multiple E&P projects. They provide intuitive discussions on the differences between financial portfolios and E&P project portfolios and proper risk-measures. However, the EPPO model is a much simplified model of reality with each project characterized solely by two parameters, NPV and CAPEX. Gustafsson's contingent project portfolio (CPP) model extends Heidenberger's work by replacing the sample path based budget constraints with the inventory type dynamic budget constraints, which allow excess cash flows can be carried to next period and earn interests. The inventory budget constraints have been used in the dynamic asset liability management (ALM) model (Consigli and Dempster, 1998, and Klaassen, 1998) and the dynamic portfolio optimization model (Korn and Korn, 2000). In such a way, all cash flows generated by one project are explicitly captured. So some investment opportunities and flexibility omitted by Heidenberger's model have been captured. Besides this advantage, the CPP model allows flexible choices of risk measures. So the CPP model is a more robust and flexible framework which would be a good starting point for use to build portfolio optimization model for exploration and production (E&P) projects.

Even though the portfolio approaches to manage risky projects have been discussed over the past two decades, there are limited reports regarding practical applications. Most previous work mainly focuses on presenting fundamental ideas with small examples. Such models avoid many practical issues, such as multivariate state trees, statistical project dependency and learning effects, and etc. Our portfolio optimization model have successfully solve problems up to (but not limited to) ten projects, over ten time periods, and with various investment decision types. From our research, we identified one main obstacle which restricts the application of the project portfolio model: the growth of the problem size. Like most multistage stochastic programming problems, once the problem size, or to be more specific, once the scenario tree size is under certain level, the MILP portfolio optimization model can be solved very quickly. Many binary decision variables can be fixed during the preprocess step. As any multi-stage stochastic programming, the scenario tree grows exponentially, which imposes even more severe numerical difficulty than the presence of discrete variables. Therefore, our solution strategy mainly focuses on incorporating different numerical and statistical methods to reduce the tree size to the level under our control.

There are several advantages of this portfolio approach over traditional isolated project assessment and selection methods. First, we can select and hold E&P projects of different risk profiles from diverse locations to hedge geological and geopolitical uncertainties in an effective and consistent way. Second, we take a global view to all the projects simultaneously and look for overall optimal strategies. Third, this dynamical approach can help us to find better strategies which traditional static and separate methods may ignore.

4.1.2 E&P project organization

To build decision trees for individual E&P projects, we need to identify the major decisions, available information, and cash flows. For this purpose, we divide the life cycle of a typical E&P project into five phases as shown in Table 4-1, each extending over one or more years. The five phases are, acquisition, appraisal and delineation, conceptual study and facility installation, production, and expansion and production. A more elaborate review about these activities is presented in §2.1. Similar phase-models have been used before, such as Meister et al. (1996), Smith and Mccardle (1999), Lund (2000), and Leffler et al. (2003).

Contingent on information available at the start of each phase, some decisions are taken at certain costs. The decisions will either continue on developing the field or abandon it. The following Table 4-1, Table 4-2, and Figure 4-1 show the E&P project's organization and the timing of all decisions and information learning.

Table 4-1 Phase descriptions of a typical exploration and production oil project

Phase #	Phase	Activities / Decisions
1	Acquisition and Preliminary exploration	♦ Collect commercial maps, conduct preliminary seismic surveys to prepare for lease auctions ♦ Estimate chance of success & commerciality ♦ Decide whether to submit a sealed bid to acquire the right to explore and develop a field ♦ Once the field is acquired, a wildcat well is drilled to verify the existence of oil deposits
2	Delineation and appraisal drilling (contingent on state "Pay")	♦ If the outcome is "Pay", the following activities continue, otherwise, the field is abandoned. ♦ Decide whether to acquire additional seismic data ♦ Based on the seismic surveys, decide whether to drill additional wildcat wells and take core samples
3	Conceptual study and facility installation	♦ Decide on how much oil gathering and processing capacity to install and which type of platforms ♦ Decide how many production wells to drill, limited by the processing capacity.
4	Production	♦ Decide how many production wells to drill, limited

		by the processing capacity.
5	Capacity expansion and continuation of production	♦ Decide whether to expand the capacity, how much to expand, and which type technology to use.

This structure is designed to provide a reasonable compromise between realism and model complexity. It is important that the decision tree representing a single project not be too large, because our goal is to develop portfolio models which contain multiple projects. However, another goal is to capture all stages of the E&P process that are important in making portfolio decisions.

Table 4-2 Phase descriptions of decisions and information

Phase #	Phase	Contingent Information	Acquired information
1	Preliminary study	♦ Public information ♦ Experts' opinions	♦ Chance of success ♦ Seismic data to estimate reservoir commerciality
	Acquisition	Acquisition decision is made based on the preliminary study: ♦ Chance of success ♦ Reservoir commerciality	
2	Delineation and appraisal	♦ Seismic surveys provide information to decide	♦ Seismic surveys estimate reservoir size and

	drilling	whether to wildcat wells ♦ Other delineation and appraisal information help to decide whether to drill more appraisal wells to delineate the reservoir size and commerciality	extractability ♦ An appraisal well is drilled to verify the existence of hydrocarbons: Pay/No Pay ♦ Core samples and seismic surveys provide estimates of commerciality (reserve size and reservoir pressure) ♦ Additional seismic surveys and appraisal refine the previous estimates
3	Conceptual study and facility installation	♦ Appraisal drillings provide information to delineate the platform type and capacity ♦ The concept study considers the flexibility to expand the capacity in the future	
4	Production		♦ Production flows provide third estimates of R and IP and second estimates of WRES
5	Capacity expansion and production	♦ The capacity expansion decision relies on the historical recovery and future potential.	

The following Figure 4-1 provides an example to show the time of the decisions and information according to the project organization in Table 4-1 and 4-2.

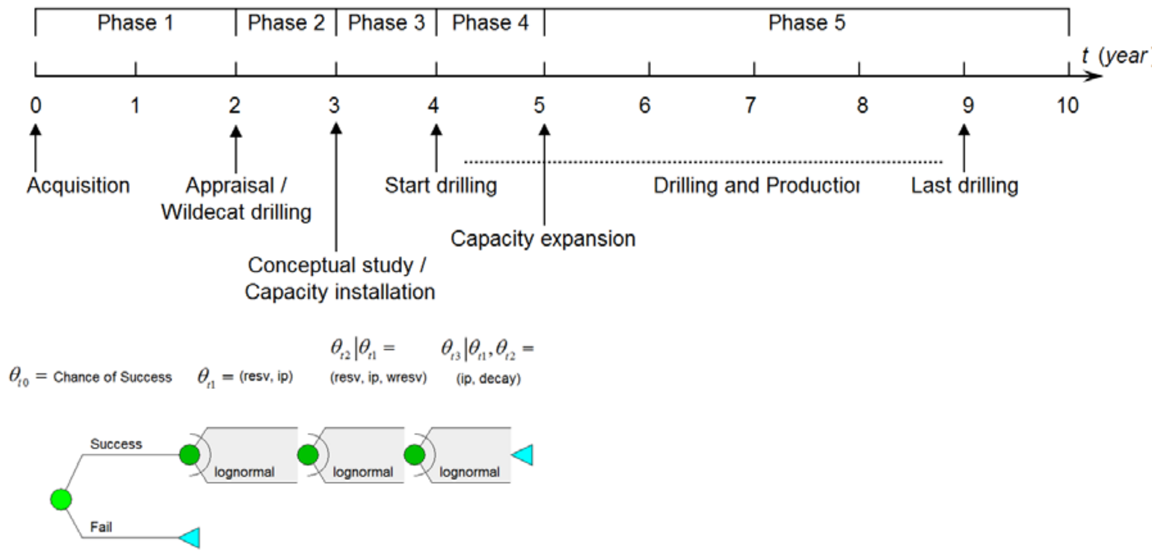


Figure 4-1 The phase model of an E&P project with proper timings of various decisions and information revelations. The top part of the figure lists decisions from left to right. The bottom part is the state tree which is aligned to the time axis as well.

In the example, the uncertainties associated with each reservoir are the chance of success, *COS*, recoverable oil reserves, *Resv*, the initial production rate of a well, *IP*, the recoverable reserves per well, *Wres*, and the decay rate, *decay*. Except for the *COS*, the distributions of other random variables are assumed to be lognormal. To incorporate statistical dependence and learning effect, we assume they are correlated in some way. We will cover this in the next chapter. Tests which provide information on the distribution of values of these quantities are made in phases 1 through 3 and prior to the drilling and production procedure.

4.2 Decision Tree Model for Individual E&P Project

As we mentioned previously, traditional asset pricing theories can not be directly used to assess E&P projects because the geological uncertainty is project-specific and can not be hedged away by trading in financial markets. So we need a more general project valuation method which can handle the incomplete market cases.

The valuation problem in incomplete markets has been extensively discussed in literature. Traditional methods are mainly based on the utility theory and the concept of certainty equivalent (CE) value. Broadie and Detemple (2004) provide a good review on this in the continuous time setting. They conclude that "… The certainty-equivalent value is a private value to the extent that it depends on the preferences of the individual holding the asset and exposed to the notraded risks. It represents the ask price for this particular individual." However, although the CE method is theoretically intuitive, it is very difficult to use in practice since general utility functions are usually non-invertible, particularly in high dimensional cases.

Unlike financial securities, whose prices are determined by public information and formed as an equilibrium result due to non-arbitrage tradings, an E&P project doesn't have a universally accepted value, or price. First, different investors may assume different utility functions over the same cash flows and thus may carry out the same project differently and hence lead to different future cash flows. Second, investors are likely to have different subjective probability assessments to the same project-specific uncertainty. Finally, investors may require different premiums to compensate for exposing to project risk of the same profile. Particularly, when investors assess a project in a portfolio perspective, the project valuation process would become even more complicated due to the budget restrictions and interactions with other on-going and potential projects.

Instead of seeking for certainty equivalent values, the approach we take to evaluate E&P projects is based on cash flow calculations and the utility maximization theory. The cash flows incurred to an E&P project are generated by various capital investment decisions and operational activities. The basis to compare different cash flows is a properly selected utility function. Utility functions are introduced to reflect the investor's time preferences and risk attitude over the random cash flows. Our goal is to find the optimal exploration and production strategies such that the resulting oil projects' random cash flows maximize the expected utility.

Each E&P project consists of a sequence of decision making and information learning activities. It may take up to thirty years to explore, develop, and deplete a reservoir. The total costs to develop an oil field could be from millions of dollars up to billions of dollars. It is a challenging task to build a solvable model which still includes crucial decisions and uncertainties occurred in the full cycle of E&P projects.

Since the E&P investment decisions are made sequentially and based on progressively revealed information, the project valuation problem can be naturally modeled as a stochastic dynamic programming (SDP) problem. The information accumulated over time can be described by a collection of increasingly refined distributions, which is called a filtration. If the utility functions are time-additive, the SDP can be formulated in the Bellman's equations. If both the state space and the decision space can be discretized, we can conveniently formulate and solve the SDP problem with decision tree analysis (DTA).

DTA provides a systematic structure to organize various information and decisions in a systematic and intuitive way. However, the solution procedure of a decision tree of practical size may be too slow to be applicable since it usually relies on an explicit exhaustive search over all branches or nodes. Motivated by the implicit enumeration of the branch-and-bound (B&B) in mixed integer programming (MIP), people have realized that it may be advantageous to use MIP techniques to model and solve decision trees since many feasible strategies which would otherwise have to be assessed on the decision tree can be ruled out by implied bounds developed during the B&B procedure. Although it is possible to expedite the DTA solution process with the implicit enumeration as well, it is still less likely to get a general purpose DTA code to beat commercial MIP solvers like CPLEX which have been specially designed and optimized to deal with large problems.

In this section, we first introduce the project organization for a typical E&P project. We then use the organization as a framework to build the decision tree model for the E&P project. The project-specific decision tree model serves as building block for our portfolio model presented in the next section.

4.2.1 Uncertainty evolution

There are different sources of uncertainties associated to an unexplored field. At the beginning, the information about the existence, size, and commerciality of underground oil deposits is scarce. Oil companies need take a sequence of exploratory activities to acquire necessary information before they decide whether to continue investing. The information is expressed in the form of probability distributions and progressively refined as long as new information comes. As more money is spent on exploration, more accurate knowledge about the reservoir performance is acquired. The improved knowledge is usually reflected by the increasingly narrowed spreads of the probability distributions.

In our E&P project model, we are particularly interested in the random parameters that are related to the prediction of field economic and operational performance. The chance of success (COS, or the chance of Pay) is the single most important factor that affects the field value and the decisions whether to acquire and explore a field. Once a successful wildcat well hit oil deposits, more information used to assess the field commerciality will be collected by various techniques at certain costs.

The commerciality of a field relies on its performance. The field performance is determined by the amount of recoverable reserve, extractability, and recovery speed. For simplicity, we choose the simplified tank model to forecast both well and field production rates over time. The simplified tank model uses an exponential decline curve to predict individual well production.

Simplified tank model

The tank model introduced in §2.2 involves lower level geological parameters and hence leads to a production forecasting model which is nonlinear in drilling decisions, particularly when wells are drilled over time. The linearization technique proposed in §2.2.4 will significantly increase the number of discrete variables, which makes it impractical to use in stochastic programming models. A simplified linear and consistent tank model will be used in our following discussions.

The simplified tank model is determined by five parameters, reservoir recoverable reserve (*Resv*), recoverable reserve per well (*WResv*), well initial production rate (IP for abbreviation, Q_0), decay rate (λ), and economic limit (Q_{el}). It is reasonable to assume that a reservoir takes a long time to reach its present inner equilibrium which maintains unanimous geological properties everywhere in the reservoir. To protect environment, there is a minimal spacing requirement imposed between any two drilling sites by regulations. We assume that spacing is distant enough such that the production in one well will not significantly impact the production behavior of other wells in the same reservoir within the E&P project lifetime. So each well's production behavior in the same reservoir can be forecasted by the same exponential decline model as follows,

$$Q(t) = Q_0 \cdot e^{-\lambda \cdot (t - t_0)}, \quad t_0 \le t \le t_m, \tag{4.2.1}$$

where t_0 denotes the drilling time and $t_m = \min\{t : Q(t) = Q_{el}\} = t_0 + \frac{1}{\lambda}\ln\frac{Q_0}{Q_{el}}$ the shut-in time. There is a consistence relationship among *WResv,* Q_0, Q_{el}, and *λ,*

$$WResv = \int_0^{t_m} Q(t)\,dt = (Q_0 - Q_{el})/\lambda. \tag{4.2.2}$$

Therefore, the maximal number of wells that can be drilled is roughly given by

$$M_w = \lfloor Resv/WResv \rfloor. \tag{4.2.3}$$

The economic limit Q_{el} is determined by the production cost, tax, and oil price. For simplicity, we usually pick a fixed economic limit for all reservoirs. So the reservoir development and production strategies will only depend on other four random geological parameters.

Evolution of random parameters of the tank model

The Figure 4-2 below shows how we represent the evolution of the random parameters of the tank model in the following E&P project model.

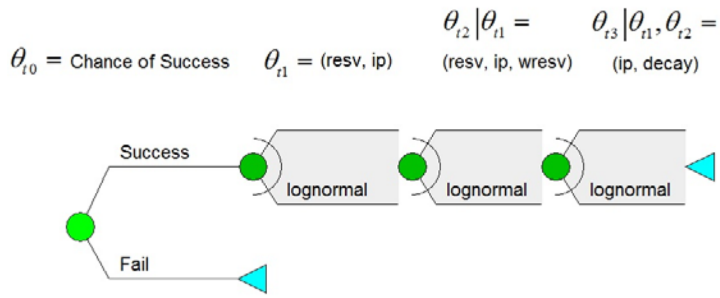


Figure 4-2 State tree of an E&P project (consistent to Figure 4-1 and Figure 4-3).

The state tree above consists of a sequence of chance nodes followed by two or more branches to reflect our present knowledge about how the uncertainties of an E&P project evolve over time. In decision tree analysis society, a round node together with its branches represents a discrete distribution of a random parameter or a discrete joint distribution of a random vector; a square node (shown in Figure 4-3) represents a decision node; and a triangular node represents a terminal end or state.

We read the state tree in Figure 4-2 from left to right, in accordance with the time axis as shown in Figure 4-1. In the sequence, the resolved information includes the chance of success, the initial estimates on reservoir size and initial well production (IP), the improved estimates on reservoir size and IP plus the first estimate on well recoverable reserve size, followed by the final updated estimates on IP and the first estimate on decay rate. Normally, in oil industry, we assume the parameters *Resv*, *WResv*, *IP*, and *Decay* follow multi-lognormal distributions. Therefore, a discretization scheme is required to generate branches for the continuous multi-lognormal distributions. We will explain this in the next chapter.

4.2.2 Decision tree model

The following Figure 4-3 is the decision tree corresponding to a typical E&P project. It incorporates the state tree (Figure 4-2) with three-point approximations to capture the underlying information structure. Like the state tree, we should read the decision tree from left to right. There are a few user-friendly software packages to build and solve decision trees. By widely-accepted conventions in the decision science society, we can conveniently identify the types of tree nodes by their shapes. The decision, chance, and terminal nodes normally have shapes of square, circle, and triangle, respectively. The software we used to build the decision trees is DPL. In the decision tree, each yellow square box represents a decision node, each green circle a chance node, and each blue triangle a terminal node. Each branch either represents one decision to be made or one potential realization (or transition) of a random parameter, depending on the type of the node from which the branch fans out.

In DPL's terminology, a tree node, either chance node or decision node, can be one of three types: symmetric, asymmetric, and mixed. A symmetric node has the same subtree succeeding on each of its branches. An asymmetric node may have different subtrees residing on its branches. A mixed node is a hybrid type of the above two types. For simplicity, the DPL only shows one realization of duplicate subtrees on a symmetric node or mixed node. For example, the decision node "Acquisition & Initial exploration" and chance node "Chance of Success" are asymmetric nodes. The decision nodes "Delineation" and "Capacity installation" are mixed nodes, each with one branch leading to a terminal node, a blue triangle. All other nodes are symmetric nodes. For more details about the decision tree, please refer to relevant textbooks or software manuals.

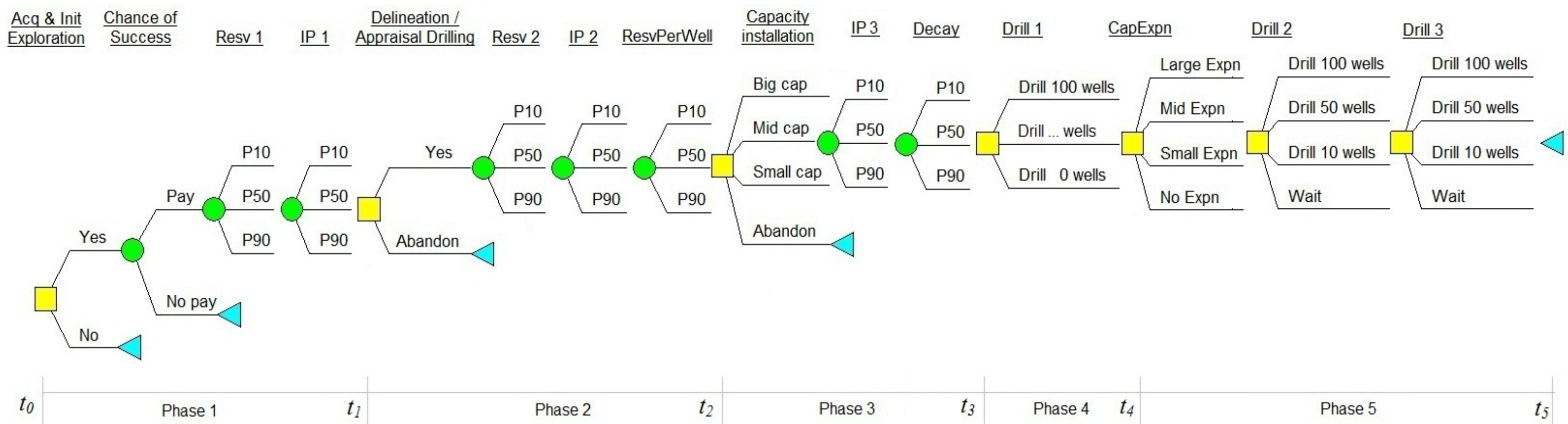


Figure 4-3 Decision tree of an E&P project

The project modeled by the decision tree in Figure 4-3 is self-explained. The first decision, whether to acquire the field and perform preliminary explorations, is the most important one. The decision must be made here and now, solely based on the present information about the future reservoir performance, expressed by the state tree in Figure 4-2, and the optimal subsequent decisions contingent on the information, expressed by the succeeding decision nodes. The entire decision tree is to take into account all the information and future decisions to assist making the first decision.

This reasoning can be recursively applied to all other decisions. At any decision node, the optimal decision is made with the assistance of the information and decisions in the subtree rooted at the node. This idea exactly reflects the Bellman's Principle of Optimality. Let's quote Dixit and Pindyck's words, "An optimal policy has the property that, whatever the initial action, the remaining choices constitute an optimal policy with respect to the subproblem starting at the state that results from the initial actions."

If the acquisition decision is "No," the project will be abandoned immediately and nothing will ever occur. Otherwise, the reservoir is acquired. The company will perform initial explorations, such as seismic surveys, wildcat drilling, and so on, to resolve the deposit existence issues and get rough estimates about the commerciality of the reservoir. More details about the E&P project decision tree will be discussed in the project portfolio model in the next section.

4.2.3 Cash flow structure

Associated with the project organization is the timing and values of incurred various cash flows, including E&P costs, revenues, interests, royalties, and taxes. The associated parameters and their timings are summarized in the following Table 4-3.

Table 4-3 Cash flow structure

Stage	Cash flow	Type	Units	Example[5]
t_0	Acquisition cost	Fixed	MM$	7
t_1	Appraisal cost	Fixed	MM$	60.60
t_2	Facility construction cost (including platform and processing units)	Mutual exclusive choice	(MMbl, MM$)	Small (25, 15) Medium (50, 35) Large (75, 50)
$t_3 \sim t_5$	Drilling cost	Fixed	MM$/well	10.70
	Production cost	Variable	$/bbl	17.90
t_4	Facility expansion cost	Mutual exclusive choice	(MMbl, MM$)	Small (20, 15) Medium (40, 32) Large (60, 50)
$t_4 \sim t_5$	Sales revenue/Oil price	Variable	$/bbl	30 (initial price)
	Royalty	Variable	%	12%
	Tax	Variable	%	30%
$t_0 \sim t_5$	Interest	Variable	%	10%

4.2.4 Cash flow calculations with royalties and taxes

Suppose we have an oil project which generates a deterministic production flow q_t, $\forall t \in T$. Given spot prices s_t, $\forall t \in T$, unit production cost, *c*, tax rate, *tax*, royalty rate, *roy*, and compounded discounting factor, *ρ*. We can compute the project NPV with

[5] The example data is excerpted from our numerical example in §4.4, corresponding to oil field 1 and Table 4-6 to Table 4-8.

$$
\begin{aligned}
NPV &= \sum_{t\in T}\left(s_t q_t \cdot (1-roy) - c\cdot q_t\right)(1-tax)\cdot e^{-\rho t} \\
&= \sum_{t\in T}\left(s_t \cdot (1-roy) - c\right)(1-tax)\cdot q_t \cdot e^{-\rho t}
\end{aligned} . \quad (4.2.4)
$$

If we have stochastic models for both the market information of spot prices and interest rates and the geological information of the production rate forecasting, we can compute the project ENPV with

$$
\begin{aligned}
ENPV &= \mathop{\mathbb{E}}_{(\tilde{s},\tilde{\rho}),\tilde{q}}\left[\sum_{t\in T}\left(\tilde{s}_t \tilde{q}_t \cdot (1-roy) - cq_t\right)(1-tax)\cdot e^{-\tilde{\rho} t}\right] \\
&= \mathop{\mathbb{E}}_{(\tilde{s},\tilde{\rho})}\left[\sum_{t\in T}\left(\tilde{s}_t \cdot (1-roy) - c\right)\cdot e^{-\tilde{\rho} t}\right]\cdot (1-tax)\cdot \mathop{\mathbb{E}}_{\tilde{q}}\left[\tilde{q}_t\right]
\end{aligned} . \quad (4.2.5)
$$

4.3 MULTISTAGE E&P PORTFOLIO OPTIMIZATION MODEL

In this section, we formulate the multistage E&P portfolio optimization model (MEPPO). We first introduce the model structure, and then discuss each component of the model in separate subsections. The MEPPO model is a multistage stochastic programming model with many discrete decision variables. Efficient solution methods are required to solve the resulting large scale MILP. In the next section, we will discuss the solution and numerical experiments of the MEPPO problem.

4.3.1 Model structure

The MEPPO model relies on a scenario tree to approximate the dynamics of all considering E&P projects. Each node in the scenario tree represents one state for all projects. We will discuss how to generate the multivariate scenario tree in Chapter 5.

The MEPPO model consists of four types of constraints as shown in Figure 4.4, the logic consistence constraints to define the decision trees, each for one project, the constraints to collect cash flows consumed or generated by the decision trees, the inventory constraints to ensure cash flow balance between any two consecutive time points along each sample path, and finally, the terminal cash position constraints to compute the terminal rewards and risk measures. The objective function measures the investors' attitude to the tradeoff between overall benefits and risks.

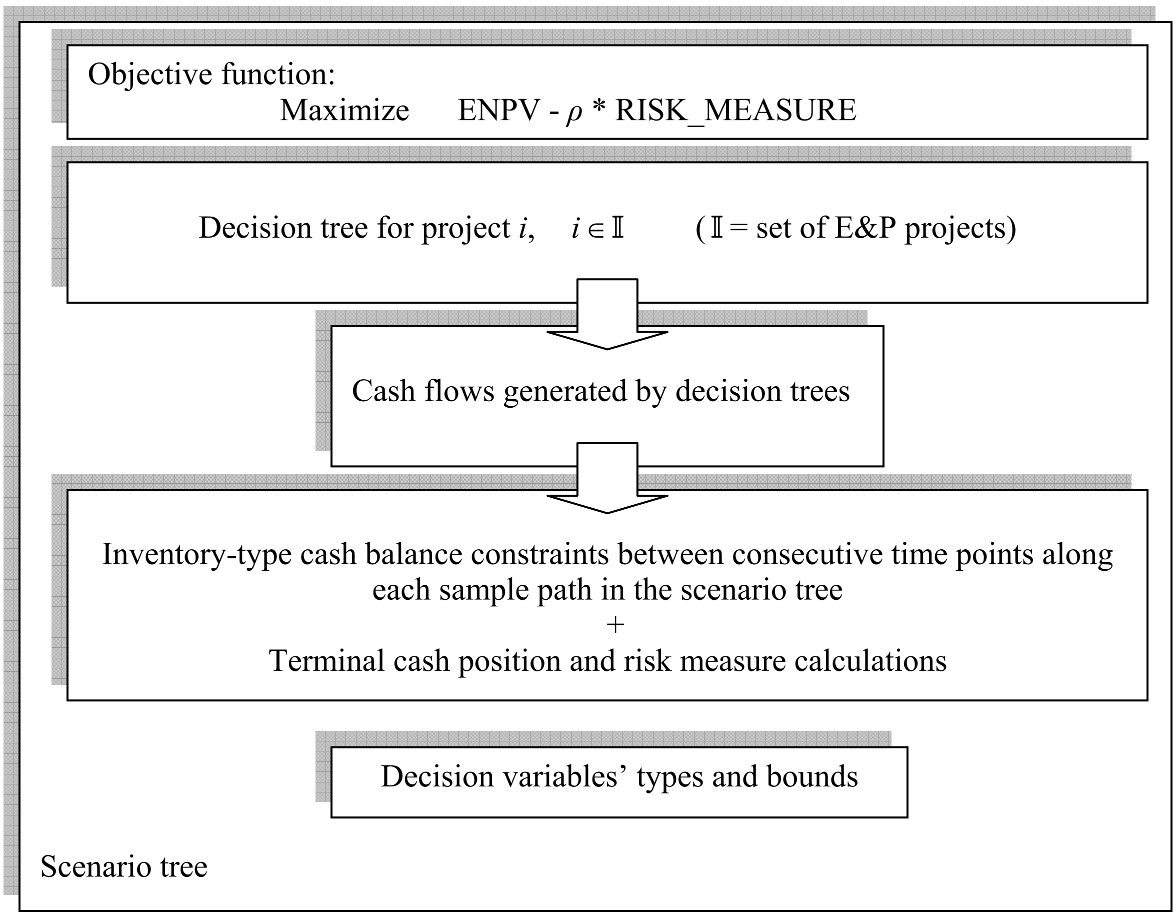


Figure 4-4 The structure of the MEPPO model

There are three types of decision variables, binary, integer, and continuous variables. The binary variables are mainly used to model the "go-no go" type decision, logic consistence decision such as "a field can be appraised only if it has been acquired," and choices among multiple production technologies. Particularly, the multiple choice decisions are declared as the special ordered set of type one (SOS1) variables to allow the solver to accelerate the branch and bound procedure. The integer variables are used to model drilling decisions, i.e., the number of wells to be drilled during each production period. The continuous variables are mainly used to calculate various cash flows.

According to the different nature of uncertainties, the scenario tree of the portfolio model consists of two subtrees. The first subtree will resolve during the first phase. Each leave node of it represents one realization of the joint binary *COS* state of all fields. The second captures the phased distributions of four continuous paremeters *resv*, *ip*, *wres*, and *decay*.

4.3.2 Modeling individual decision trees

We now take the approach proposed by Gustafsson and Salo (2005) to formulate the E&P projects as decision trees with mixed integer programming (MIP) techniques. Let us start with some commonly used notations.

Sets and indices

$\mathbb{I}$: the set of oil projects, each for one oil field, subscripted by *i*.

$\mathbb{T}$: the set of decision stages, $\mathbb{T} = \{0, \cdots, T\}$, subscripted by *t*.

$\mathbb{PH}$: the set of phases of the scenario tree, $\mathbb{PH} = \{0, \cdots, 3\}$, subscripted by *ph*.

$\mathbb{L}$: the set of platforms with different production capacities, subscripted by *l*.

$\mathbb{K}$: the set of expansion technologies with different capacities, subscripted by *k*.

$\mathbb{S}$: the set of nodes on the *COS* subtree of the scenario tree, superscripted by *s*.

$\mathbb{N}$: the set of nodes on the second subtree of the scenario tree, superscripted by *n*.

$\mathbb{S}\times\mathbb{N}$: the set of nodes of the composite scenario tree, superscripted by (s,n).

$\mathbb{N}_{ph}$: the phase-*ph* node set, $\mathbb{N} = \bigcup_{ph\in\mathbb{PH}} \mathbb{N}_{ph}$, subscripted by n_{ph} accordingly.

Mappings

$a:\mathbb{S}\times\mathbb{N}\to\mathbb{S}\times\mathbb{N}$ returns the immediate predecessor of the operand.

$a_t:\mathbb{S}\times\mathbb{N}\to\mathbb{S}\times\mathbb{N}$ returns the stage-*t* predecessor of the operand along the path.

$t:\mathbb{PH}\to\mathbb{T}$ returns the ending time stage of the operand (a scenario phase).

$ph:\mathbb{T}\to\mathbb{PH}$ returns the scenario phase of the operand (a time stage).

Given a stage-*t* node (*s, n*), $a_\tau(s,n), \tau<t$, returns its stage-*τ* predecessor along the sample path backwardly in the scenario tree. Particularly, if no new information is resolved between stage-*τ* and stage-*t*, $a_\tau(s,n)$ returns the operand itself, i.e. (*s, n*). We use the following simple example in Figure 4-5 to explain those notations.

Figure 4-5 shows how we model the E&P portfolio uncertainty with a composite scenario tree, which consists of two independent subtrees. A portfolio state (s,n) includes two components, *s* and *n*. The state $s\in\mathbb{S}$ comes from the bottom subtree, called subtree-1, which represents one realization of the binary "Pay/No pay" combinatorial state of all oil fields. The state $n\in\mathbb{N}$ comes from the top subtree, called subtree-2, which represents how the tank model parameters of each field evolve over time. The subtree-1 resolves completely during the first phase. In the mean time, the first phase of the subtree-2 resolves by wildcat drilling. The subtree-2 resolves completely after the third phase at t_3. For any three states $n_1\in\mathbb{N}_1, n_2\in\mathbb{N}_2, n_3\in\mathbb{N}_3$ in the subtree-2, if they are on the same sample path, by the recursive definition of the mapping $a(.)$, we have $a\left(a\left(a(s,n_3)\right)\right) = a\left(a(s,n_2)\right) = a(s,n_1) = (s_0,n_0)$. E.g. $a(2,9)=(2,4)$, $a(2,4)=(2,1)$, $a(2,1)=(2,0)$, $a_2(2,9)=(2,1)$, and $a_1(2,9)=(2,0)$.

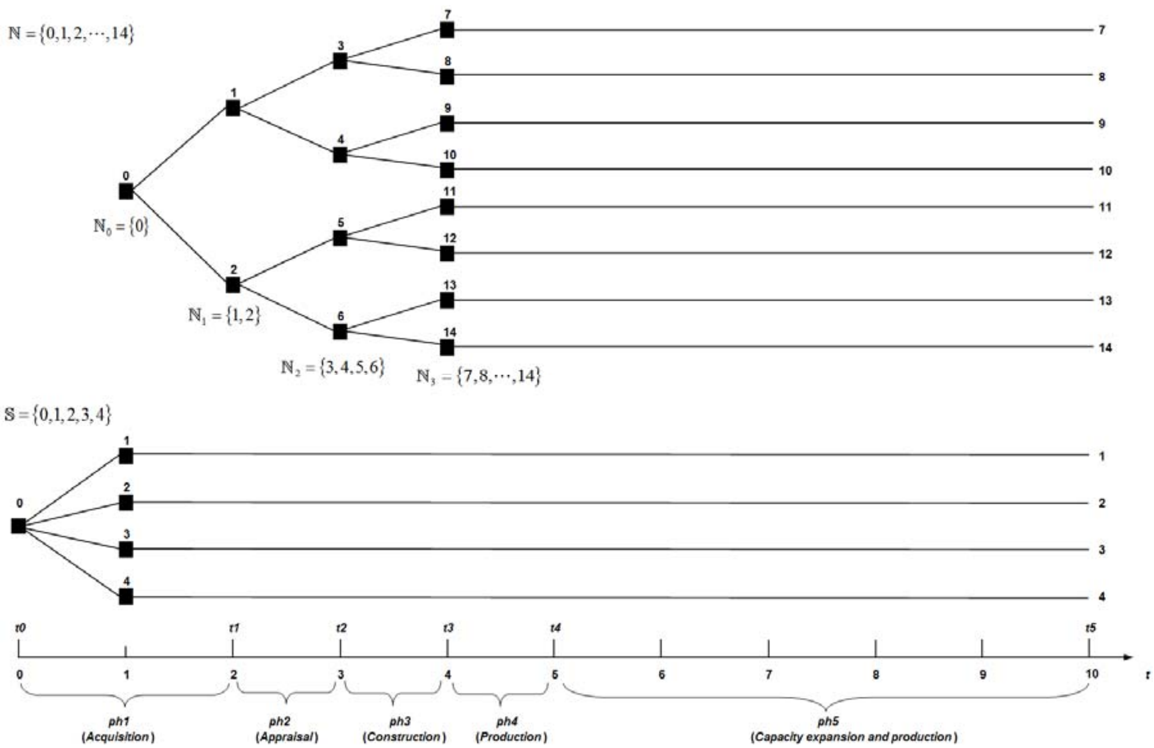


Figure 4-5 Composite scenario tree

Naming conventions

We use different naming styles as described below to differentiate parameters from decision variables, discrete variables from continuous variables, and scenario related indices from non-scenario related indices.

Uppercase and lowercase: Parameters are denoted with an uppercased initial letter and decision variables with all lowercased letters. For example, $x\_acqr_{i,t}^{s_0,n_0}$ is a decision variable and Acq_Cost_i is a parameter.

Prefixes: All discrete variables are prefixed by "*x*_" and all continuous variables are prefixed with "*y*_." For example, $x\_acqr_{i,t}^{s_0,n_0}$ is a binary variable for the decision whether to acquire reservoir *i* at time *t* in the root node state (s_0, n_0), while $y\_rev_{i,t}^{s,n}$ is a continuous variable modeling the sales revenue from field *i* at time *t* in the state (*s, n*).

Subscripts and superscripts: We use superscripts s and n to denote a state of the composite scenario tree for a decision variable or parameter. We use subscripts to denote indices of non-scenario indices. For example, $y\_rev_{i,t}^{s,n}$ explained above.

Parameters

Acq_Cost_i : the acquisition cost of field *i*.

$Appr\_Cost_i$: the appraisal cost of field *i*.

$DrillCost_i$: the per-well drilling cost at field *i*.

$ProdCost_i$: the unit production cost at field *i*.

$Cap1_l$: the initial capacity for the capacity type-*l*.

$CapCost1_l$: the capacity installation cost for the capacity type-*l*.

$Cap2_k$: the expanded capacity for the capacity expansion plan type-*k*.

$Cap2Cost_k$: the capacity installation cost for the capacity expansion plan type-*k*.

$\bar{Q}_i^{n_3}$: per-well initial production rate at field *i* in state n_3.

$\lambda_i^{n_3}$: the annual well production decay rate at field *i* in state n_3.

$M_i^{n_3}$: the maximal number of wells that can be drilled at field *i* in state n_3.

$Pr1^s$: the probability of the subtree-1's state $s \in \mathbb{S}$.

$Pr2^n$: the probability of the subtree-2's state $n \in \mathbb{N}$.

$Prob^{s,n} = Pr1^s \cdot Pr2^n$: the joint probability of the composite state $(s,n) \in \mathbb{S} \times \mathbb{N}$.

Cash flow generated by the decision tree

$y\_capex_{i,t}^{s,n}$: the CAPEX consumed by field *i* at time *t* in state $(s,n) \in \mathbb{S} \times \mathbb{N}_{ph(t)}$.

$y\_opex_{i,t}^{s,n_3}$: the OPEX consumed by field *i* at time $t \geq t_3$ in state (s,n_3).

$y\_rev_{i,t}^{s,n_3}$: the revenue generated by field *i* at time *t* in state (s,n_3).

We can start present the model by defining decision variables and write down relevant constraints in the sequence of how oil fields are developed.

Acquisition and initial exploration at $t = t_0 = 0$

We define a binary decision variable $x\_acqr_{i,t}^{s_0,n_0}$ to model the acquisition decision and preliminary exploration,

$$x\_acqr_{i,t}^{s_0,n_0} = \begin{cases} 1, & \text{if we acquire field } i \text{ at stage } t \text{ in the root state } (s_0, n_0). \\ 0, & \text{otherwise.} \end{cases}$$

Once the acquisition decision is made, one or more test wells will be drilled to verify the existence of oil deposits. The test result will help to make the succeeding appraisal decision. So the acquisition cost Acq_Cost_i should include the cost to drill the test well. The binary variable $x\_acqr_{i,t}^{s_0,n_0}$ determines whether the fixed cost Acq_Cost_i is incurred. The relevant capital expenditure is calculated with

$$y\_capex_{i,t_0}^{s_0,n_0} = Acq\_Cost_i \cdot x\_acqr_{i,t_0}^{s_0,n_0}, \quad \forall i, \tag{4.3.1}$$

where (s_0, n_0) denotes the root node of the scenario tree.

Delineation and appraisal exploration at $t = t_1$

The outcome of the preliminary test can be either "Pay" with probability equal to the chance of success, *COS,* or "No pay" with probability of 1-*COS*. If the former case happens, we have the option to decide whether to continue the appraisal and delineation. Otherwise, the field will be abandoned permanently and the acquisition cost is never recovered. To capture this dependence, we introduce a new parameter $Pay: \mathbb{S} \times \mathbb{I} \rightarrow \{0,1\}$ to indicate the realization of the "Pay/No pay" state for each field.

$$Pay(s,i) = \begin{cases} 1, \text{if the state is "Pay" at field } i \text{ in state } s. \\ 0, \text{otherwise.} \end{cases}$$

Let binary decision variable x_appr_{i,t_1}^{s,n_1} to denote the binary decision whether to continue the appraisal investment at field *i*.

$$x\_appr_{i,t_1}^{s,n_1} = \begin{cases} 1, & \text{if field } i \text{ is appraised at } t_1 \text{ in state } (s,n_1). \\ 0, & \text{otherwise.} \end{cases}$$

So we have the following logic constraints,

$$x\_appr_{i,t_1}^{s,n_1} \le Pay(s,i) \cdot x\_acqr_{i,t_0}^{s_0,n_0}, \quad \forall (s,n_1) \in \mathbb{S} \times \mathbb{N}_1, \tag{4.3.2}$$

such that $(s_0,n_0) = a(s,n_1)$. Therefore, the appraisal decision is allowed only if the field is acquired and oil deposits are identified.

The appraisal and delineation cost is calculated with

$$y\_capex_{i,t_1}^{s,n_1} = Appr\_Cost_i \cdot x\_appr_{i,t_1}^{s,n_1}, \quad \forall i,s,n_1. \tag{4.3.3}$$

Platform selection and capacity installation at $t = t_2$

After explorations conducted in the acquisition and appraisal phases, we have collected necessary information to estimate the reserve size and perform conceptual study. The purpose of the conceptual study is to assist the decision making to choose a proper hydrocarbon lift technology and its annual production capacity.

We define binary variables $x\_cap1_{i,l}^{s,n_2}$ and continuous variables $y\_cap1_i^{s,n_2}$ to model the capacity installation decision and capture the installed capacity, respectively,

$$x\_cap1_{i,l}^{s,n_2} = \begin{cases} 1\ , & \text{if the type-}l\text{ platform is installed at reservoir } i \text{ in state } (s,n_2), \\ 0, & \text{otherwise,} \end{cases}$$

$y\_cap1_i^{s,n_2}$: the capacity installed before production in state (s,n_2) at reservoir *i*.

For simplicity, we ignore the exact type of available platforms and equipped technology. We simply differentiate them by their processing capacity and installation cost and quote them as a concept set $\mathbb{L}$. Each concept l in $\mathbb{L}$ is characterized by a unique pair of annual processing capacity, $Cap1_l$, and one-time fixed installation cost, $CapCost1_l$. Each concept allows expansion at a later time according to a set $\mathbb{K}$ of capacity expansion plans.

We can construct production infrastructure and install capacity to reservoir *i* only if we have finished appraisals. So we have the following logic relationship,

$$\sum\nolimits_{l\in L} x\_cap1_{i,t_2,l}^{s,n_2} \le x\_appr_{i,t_1}^{a(s,n_2)},\ \ \forall i,s,n_2 \tag{4.3.4}$$

Clearly, at most one variable of the left hand side of (4.3.4) can take a nonzero value. This happens only if the right hand side takes nonzero value, i.e. the reservoir has been acquired appraised. The installed capacity at reservoir *i* in state (s,n_2) is captured by

$$y\_cap1_i^{s,n_2} = \sum\nolimits_{l\in L} Cap1_l \cdot x\_cap1_{i,l}^{s,n_2}\ ,\ \ \forall i,s,n_2 \tag{4.3.5}$$

and the incurred capacity installation cost is

$$y\_capex_{i,t_2}^{s,n_2} = \sum_{l \in L} Cap1Cost_l \cdot x\_cap1_{i,l_2}^{s,n_2}, \ \ \forall i, s, n_2. \qquad (4.3.6)$$

We assume that once a capacity is installed its processing capacity will be maintained and available onwards. So we can suppress the time index of $y\_cap1_i^{s,n_2}$.

Drilling decisions between stage t_3 and stage $t_5 - 1$

The drilling decision and associated production plans are respectively modeled by two decision variables, the nonnegative integer variable $x\_drill_{i,t}^{s,n_3} \in \mathbb{Z}_+ \cup \{0\}$ and the nonnegative continuous variable $y\_prod_{i,t,\tau}^{s,n_3} \geq 0$ for $\tau \geq t$, where

$x\_drill_{i,t}^{s,n_3}$ = the number of wells to drill at stage *t* at reservoir *i* in state (s, n_3),

$y\_prod_{i,t,\tau}^{s,n_3}$ = the annual production during period-τ ($\geq t$) of the wells that are drilled at stage *t* at reservoir *i* in state (s, n_3).

By definition, the list $\left\{y\_prod_{i,t,t}^{s,n_3}, y\_prod_{i,t,t+1}^{s,n_3}, \cdots, y\_prod_{i,t,\tau}^{s,n_3}\right\}$ forecasts the annual productions over the duration of length $\tau - t$ by wells drilled at time *t*.

Drilling is allowed only if the platform has been stalled. We model this logic dependence with

$$\sum_{t=t_3}^{T-1} x\_drill_{i,t}^{s,n_3} \leq M_i^{s,n_3} \cdot \sum_{l \in L} x\_cap1_{i,l}^{a(s,n_3)}, \ \ \forall i, s, n_3, \qquad (4.3.7)$$

where M_i^{s,n_3} is a so-called "big-M" constant. The inequality applies an upper bound to the total number of wells that can be drilled in reservoir *i* in state (s,n_3). The bound M_i^{s,n_3} is estimated according to (4.2.3), by dividing the total reserve at reservoir *i* with the reserve recovered by per well in state (s,n_3). If any capacity is installed in reservoir *i*, we have $\sum_{l\in L} x\_cap1_{i,l}^{a(s,n_3)} = 1$ and the RHS of (4.3.7) equals to M_i^{s,n_3}. In this case, the drilling decisions are enabled and the allowable total number of production wells is capped by M_i^{s,n_3}. Otherwise, the RHS of (4.3.7) equals zero and the drilling decisions are forbidden along this sample path determined by the terminal state (s,n_3).

The drilling cost at stage *t*, such that $t_3 \le t < t_4$ or $t_4 < t < T$, is calculated by

$$y\_capex_{i,t}^{s,n_3} = DrillCost_i \cdot x\_drill_{i,t}^{s,n_3}, \quad \forall i,s,n_3. \tag{4.3.8}$$

Notice, the equation (4.3.8) excludes the drilling cost at stage t_4. Since there is an additional capital expenditure term other than the drilling cost, we leave the CAPEX calculation at stage t_4 alone for a special treatment in equation (4.3.13).

Annual production starts at stage t_3 and last up to the final stage $t_5 = T$

The geological uncertainties are significantly reduced when to make drilling and production decisions. So we can use the deterministic tank model to forecast production rates. Given revealed parameters, per-well initial production (IP) rate, $\bar{Q}_i^{n_3}$, and decay rate, $\lambda_i^{n_3}$, we can forecast the production profile for any well by the exponential decline curve, $Q_{i,\Delta t}^{n_3} = \bar{Q}_i^{n_3} \cdot \exp\{-\lambda_i^{n_3} \cdot \Delta t\}$, where Δt denotes the elapsed production time after drilling. We define a variable $y\_prod_{i,\tau,t}^{s,n_3}$ to capture the production rate at stage *t* that is contributed by wells drilled at some earlier time τ. So we have,

$$y\_prod_{i,\tau,t}^{s,n_3} = Q_{i,t-\tau}^{n_3} \cdot x\_drill_{i,\tau}^{s,n_3}, \quad \forall i, s, n_3, t_3 \le \tau \le t \le T. \tag{4.3.9}$$

We introduce a nonnegative continuous variable $y\_total_t^{s,n_3}$ to denote the total production rate during period *t,*

$$y\_total_{i,t}^{s,n_3} = \sum_{\tau=t_3}^{t} y\_prod_{i,\tau,t}^{s,n_3}, \quad \forall i, s, n_3, t_3 \le t \le T. \tag{4.3.10}$$

The production cost is captured by

$$y\_opex_{i,t}^{s,n_3} = ProdCost_i \cdot y\_total_{i,t}^{s,n_3} \ \forall i, s, n_3, t_3 \le t \le T. \tag{4.3.11}$$

Capacity expansion and capacity constraints at $t=t_4$

The concept study allows expanding the capacity at a later time, depending on new information about the reservoir performance and budget availability. Expansion plans are denoted by the set $\mathbb{K}$. We define a binary decision variable for each $k \in K$,

$$x\_cap2_{i,k}^{s,n_3} = \begin{cases} 1, & \text{if the expansion plan } k \text{ is taken at reservoir } i \text{ in state } (s, n_3), \\ 0, & \text{otherwise.} \end{cases}$$

For simplicity, we assume the capacity expansion will be performed at stage t_4, one period after the first drilling. We can suppress the time index of the expansion decision as well. The capacity expansion is allowed only if the primary capacity has been installed, so,

$$\sum_{k \in K} x\_cap2_{i,k}^{s,n_3} \le \sum_{l \in L} x\_cap1_{i,l}^{a(s,n_3)}, \forall i, s, n_3. \tag{4.3.12}$$

Each expansion plan k is characterized by the pair of expansion cost $Cap2Cost_k$ and expanded capacity $Cap2_k$. So the total capital expenditure at time t_4, which has been purposely omitted in (4.3.8), equals to the sum of drilling costs and expansion costs,

$$y\_capex_{i,t_4}^{s,n_3} = \sum_{k \in K} \left(CapCost2_k \cdot x\_cap2_{i,k}^{s,n_3} \right) + DrillCost_i \cdot x\_drill_{i,t_4}^{s,n_3}, \quad \forall i, s, n_3. \tag{4.3.13}$$

Therefore, the additional expanded capacity is given by

$$y\_cap2_i^{s,n_3} = \sum_{k \in K} Cap2_k \cdot x\_cap2_{i,k}^{s,n_3}, \ \ \forall i, s, n_3. \tag{4.3.14}$$

Once we have setup the formulae to compute the primary capacity and the secondary capacity with $y\_cap1_i^{a(s,n_3)}$ and $y\_cap2_i^{s,n_3}$, we can model the capacity constraints for the production plans. Notice, before the capacity expansion at time t_4, the annual production rate in reservoir i can not exceed the capacity $y\_cap1_i^{s,n_3}$. After the secondary capacity has setup, the capacity is raised up to $y\_cap1_i^{a(s,n_3)} + y\_cap2_i^{s,n_3}$, where $y\_cap2_i^{s,n_3}$ could be zero if none of expansion plans is taken.

$$y\_total_{i,t}^{s,n_3} \le y\_cap1_i^{a(s,n_3)}, \ \ \forall i, s, n_3, t_3 \le t \le t_4. \tag{4.3.15}$$

$$y\_total_{i,t}^{s,n_3} \le y\_cap1_i^{a(s,n_3)} + y\_cap2_i^{s,n_3}, \ \ \forall i, s, n_3, t_4 < t \le T \tag{4.3.16}$$

We assume annual production during any production year can be sold out on the spot market at the end of the year. Since the forward curve $\left\{ F_{0,t} \right\}_{t=t_0}^{T}$ gives a rough estimation to future spot prices over a strip of time points, we can use the forward price to estimate the sales revenue. The expected revenue at the stage t is

$$y\_rev_{i,t}^{s,n_3} = F_{0,t} \cdot y\_total_{i,t-1}^{s,n_3}, \quad \forall i,s,n_3,t_3+1 \le t \le T-1. \qquad (4.3.17)$$

We assume the production during the last period can be sold at the end of the last period. The resulting revenue is a crude way to take into account the salvage value of the project when it is prematurely terminated. So we have

$$y\_rev_{i,T}^{s,n_3} = F_{0,T} \cdot \left(y\_total_{i,T-1}^{s,n_3} + y\_total_{i,T}^{s,n_3} \right), \quad \forall i,s,n_3. \qquad (4.3.18)$$

The rejection and abandonment of a project

An E&P project may be rejected at the beginning or prematurely abandoned during exploration and production, depending on the balance between the overall cost to complete the project and its expected future rewards upon completion.

Rejecting a project is enforced by setting the acquisition decision to be zero at the root node, i.e., $x\_acqr_{i,t_0}^{s_0,n_0} = 0$. All subsequent decision variables will be disabled by corresponding logic consistence constraints.

In our model, we have introduced various decision variables to denote whether to continue or abandon an oil field at each phase. The meaning should be clear by which value a decision variable takes.

4.3.3 Modeling dynamic budget constraints

We have setup all decision variables and constraints necessary to model the project-specific decision trees in §4.3.2. Besides, we also have introduced variables to collect the cash flows consumed and generated by each individual project. The cash flows will be used to form dynamic budget constraints for our multistage E&P project portfolio optimization (MEPPO) model.

The dynamic budget constraints can be formulated in two different ways. The first way is just like the constraints (3.4.20),,which model path-wise budget constraints to aggregate all capital expenditures along each sample path to conform to the initial budget. This way ignores the opportunity to reinvest the generated revenues. The second way enforces a capital budgeting constraint for each time period and in each state. The available budget at the beginning of each period comes from two sources. The first source is the cash inventory carried from the previous period's cash surplus and accrued interest. The second is from selling crude oil produced from the previous period. These two budget sources fund the current period's capital investment decisions and operational activities. Any remained cash surplus will be carried as inventory over to the next period. If we are forbidden to reinvest the revenues in the second way, the two types of dynamic budget constraints are equivalent to each other.

In the following development of MEPPO model, we will take the second inventory-style approach to model the dynamic budget constraints.

We define a new nonnegative decision variable $y\_inv_t^{s,n} \geq 0$ to capture the inventory cash position.

$y\_inv_t^{s,n}$: the cash surplus after the stage-*t* investment decisions are made in state (*s, n*), which is carried over to the next period as an inventory.

The non-negativity ensures the any budget allocation decision in any state should never exceed the available budget during any period. The inventory cash position remained after budget allocation at the beginning of a period together with accrued interest will serve as the available budget at the beginning of the next period.

Dynamic capital budgeting constraints for $t = t_0$

The very first capital budgeting constraint models the decision how to allocate an initial budget B_0 to acquire some or all oil projects, which is

$$\sum_i \left(y\_capex_{i,t_0}^{s_0,n_0} \right) + y\_inv_{t_0}^{s_0,n_0} = B_0 . \tag{4.3.19}$$

Plugging the equation (4.3.1) into the LHS of (4.3.19) gives us

$$\sum_i \left(Acq\_Cost_i \cdot x\_acqr_{i,t_0}^{s_0,n_0} \right) + y\_inv_{t_0}^{s_0,n_0} = B_0 .$$

Dynamic capital budgeting constraints for $t_0 + 1 \le t < t_3$

During the period $t_0 + 1 \le t \le t_3$, all investment decisions are capital intensive expenditure decisions. No revenue is generated because no production well is drilled and completed. So the available budget at the beginning of each period only comes from the inventory cash account held from the previous period.

We have assumed that all capital investment decisions are made at the beginning of a period and any remaining fund will be deposited in a bank account to earning interests r_0 over the present period. We can write down the dynamic budget constraints in the inventory form as,

$$\sum_i y\_capex_{i,t}^{s,n} + y\_inv_t^{s,n} = B_t + \left(1+r_0\right)\cdot y\_inv_{t-1}^{a_{t-1}(s,n)},$$
$$\forall n \in \mathbb{N}_{ph(t)}, t_0 + 1 \le t < t_3. \qquad (4.3.20)$$

Dynamic capital budgeting constraints for $t_3 \le t \le T-1$

Drilling decisions are enabled only after some platform and necessary facility have been constructed. Each well incurs a fixed drilling cost, a stream of annual operating or production costs, and a stream of annual sales revenues. In the preceding section, we have setup equations to calculate them with $y\_capex_{i,t}^{s}$, $y\_opex_{i,t}^{s}$, and $y\_rev_{i,t}^{s}$, respectively. After we take into account the royalty rate r_{royal} and the tax rate r_{tax}, we get the dynamic budget constraint as follows,

$$\sum_i y\_capex_{i,t}^{s,n_3} + \left(1-r_{tax}\right)\cdot\sum_i y\_opex_{i,t}^{s,n_3} + y\_inv_t^{s,n_3} =$$
$$\left(1+r_0\right)\cdot y\_inv_{t-1}^{a_{t-1}(s,n_3)} + \left(1-r_{royal}\right)\cdot\left(1-r_{tax}\right)\cdot\sum_i y\_rev_{i,t}^{s,n_3},$$
$$\forall s, n_3 \in \mathbb{N}_3, t_3 \le t \le T-1. \qquad (4.3.21)$$

Portfolio value at the terminal stage $t = T$

For simplicity, we assume all reservoirs will be abandoned at no cost after the last decision period and no salvage value. Such assumptions are crude but reasonable since the investment horizon of an E&P project usually extends beyond ten years and the abandonment cost and salvage value won't impact the exploratory decisions too much due to the discounting effects. So the incurred cash flows other than the inventories are the production costs $y\_opex_{i,T}^{s,n_3}$ during the last period and the revenue $y\_rev_{i,T}^{s,n_3}$ by selling the produced petroleum. Taking into account the tax and royalty terms, we get the cash flow balance equation for the last period as follows,

$$
\begin{aligned}
\left(1-r_{tax}\right)\cdot\sum_i y\_opex_{i,T}^{s,n_3} + y\_inv_T^{s,n_3} &= \left(1+r_0\right)\cdot y\_inv_{T-1}^{s,n_3} \\
&\quad +\left(1-r_{royal}\right)\cdot\left(1-r_{tax}\right)\cdot\sum_i y\_rev_{i,T}^{s,n_3}, \\
&\qquad \forall s, n_3 . \qquad (4.3.22)
\end{aligned}
$$

Rewriting (4.22) gives us a formula to calculate the terminal net cash position,

$$
\begin{aligned}
y\_inv_T^{s,n_3} &= \left(1+r_0\right)\cdot y\_inv_{T-1}^{s,n_3} \\
&\quad +\left(1-r_{tax}\right)\cdot\left[\left(1-r_{royal}\right)\cdot\sum_i y\_rev_{i,T}^{s,n_3} - \sum_i y\_opex_{i,T}^{s,n_3}\right].
\end{aligned}
\qquad (4.3.23)
$$

In a similar way, we can rewrite all the cash flow balance equations (4.3.19) ~ (4.3.21) in the form of (4.3.23) to compute the inventory cash position at each stage. Notice the recursive structure of those equations via the inventory variables and we can see the above equation actually calculates the net (future) value of the portfolio developed over the investment horizon, excluding royalties and taxes.

The recursive structure can be used to prove the equivalence between the two ways to formulate the dynamic capital budgeting constraints, which we have mentioned at the beginning of this section.

4.3.4 Risk measure and objective function

We have established decision variables and constraints to model the individual E&P projects, dynamic budget constraints, and cash flows surrounding the E&P portfolio. They work together to define the feasible space of dynamic project portfolio management strategies. From the strategy space, we want to search for the best project mix and associated optimal field development plans. Our objectives are to maximize the E&P portfolio value while minimizing the investment risk. To quantify the trade-off between the two conflicting objectives, we need choose proper measures for them and an objective function to integrate them. The objective function should capture the ours preferences on how to trade off total rewards against overall risk.

There are well-established theories and models about how to choose those measures for a variety of asset classes. We use the same measures and mean-risk objective function as what Ball and Savage (1999) and Gustafsson and Salo (2005) have proposed to deal with project portfolios. To be specific, we choose the expected portfolio value at the final stage as the reward measure and the expected downside deviation of the portfolio value as the risk measure. In this way, we can formulate the MEPPO problem as a linear model. More interesting discussions about the reward and risk measures please refer to Zenios (2007).

We introduce the following notations to capture the final stage's portfolio expected value and deviations w.r.t this value.

Parameter

$\rho \geq 0$: the penalty for per unit of the expected downside deviation.

Variables

$z^*(\rho)$: the optimal objective function for the chosen $\rho \geq 0$.

y_expval: the expected terminal value of the E&P portfolio.

$y\_expdev \geq 0$: to be the expected downside deviation.

$u^{s,n_3} \geq 0$: the upside deviation of the portfolio value in state (s, n_3).

$v^{s,n_3} \geq 0$: the downside deviation of the portfolio value in state (s, n_3).

All deviations are measured in absolute values. We can integrate the conflicting objectives with the following parameterized bi-linear objective function, for $\rho \geq 0$,

$$z^*(\rho) = \text{Max} \quad y\_expval - \rho \cdot y\_expdev, \tag{4.3.24}$$

where the nonnegative parameter ρ determines the weight to penalize the risk measure, i.e., the expected downside deviation. To ensure u^{s,n_3} and v^{s,n_3} correctly capture the deviations, we need the following three formulae,

$$y\_expval = \sum_{(s,n_3)} Prob^{s,n_3} \cdot y\_inv_T^{s,n_3}, \tag{4.3.25}$$

$$y\_inv_T^{s,n_3} - y\_expval - u^{s,n_3} + v^{s,n_3} = 0, \quad \forall n_3 \in \mathbb{N}_3, \tag{4.3.26}$$

$$y\_expdev = \sum_{(s,n_3)} Prob^{s,n_3} \cdot v^{s,n_3}. \tag{4.3.27}$$

It is easy to verify with contradiction that, for $\rho > 0$, any optimal solution of the MEPPO problem defined by (4.3.1) ~ (4.3.27) must ensure that at least one of u^{s,n_3} and v^{s,n_3} should be zero in any terminal state (s, n_3). Since otherwise, both u^{s,n_3} and v^{s,n_3} are strictly positive and we always can improve the optimal objective function value by a strictly nonzero amount $\rho \cdot \min\{u^{s,n_3}, v^{s,n_3}\}$ by reducing both u^{s,n_3} and v^{s,n_3} by the same amount of $\min\{u^{s,n_3}, v^{s,n_3}\} > 0$. Obviously, this contracts to the assumption that the solution is optimal before we modify it.

It is easier to see so if we rewrite (4.3.26) as $u^{s,n_3} - v^{s,n_3} = y\_inv_T^{s,n_3} - y\_expval$. We can see that u^{s,n_3} and v^{s,n_3} do capture the proper deviation with respect to the expected terminal portfolio value.

4.4 SOLUTION METHODS AND NUMERICAL EXPERIMENTS

Like other multistage stochastic programming problems with integer variables, the MEPPO model is very difficult to solve since the number of discrete variables grows proportionally to the size of the scenario tree. It is common for the GAMS/CPLEX solver with default settings to take tens of minutes to just find a feasible integer solution for small instances with five projects, ten periods, and a few hundreds of terminal states. More often than not, a naive GAMS/CPLEX call returns nothing helpful after an hour of running. Efficient algorithms must be designed to expedite the solution procedure.

Our numerical experiments identify that the major computational challenges come from three aspects, the size of the scenario tree, the integrality of the drilling decisions, and the number of projects. For a fixed number of projects, if we can control the growth of the tree size and relax the integral drilling decisions, we can readily solve instances up to ten projects, ten time periods, and thousands of terminal scenarios. For simplicity, we assume the scenario tree has been generated with proper size and statistical quality. In this section, we will solely focus on the development of efficient solution methods. In Chapter 5, we will pay attention to discuss how to generate scenario trees with controllable growth rates.

4.4.1 Solution methods

Since the majority of discrete variables are the drilling decisions, a natural thought would be to relax the integrality of the drilling decisions, hold other binary variables unchanged, solve the relaxed MIP problem (called RMIP), and round off the continuous drilling solutions to its closest integral values. We combine the binary capital decisions of the RMIP and its rounding-off integral drilling decisions together and call them as the rounding-off solution of the RMIP. However, the rounding-off solutions are rarely feasible and sometimes they are too far to be feasible. The main reasons for the infeasibility are that all drilling decisions are involved in the inventory balance equalities and there are a great number of such equalities. The only continuous variables in the inventory constraints are also exactly determined by previous exploration and drilling decisions. Particularly, the cash flows incurred by each drilling decision are fixed numbers appearing in a few inventory balance constraints. All of those factors make the inventory balance equalities as so-called hard constraints.

We develop three optimization based heuristic methods to deal with the MEPPO problem. The first two methods try to recover an integer feasible solution from the solution of the RMIP. The third method attempts to expedite the other solution procedures when the budget is too tight. We call the optimization problems underlying the three heuristic methods as RNDOFF, DCMP, and NPSK, respectively.

To be specific, the first heuristic method solves an inverse problem (RNDOFF) to find the minimal required budget to finance the rounding-off solution of the RMIP. The result answers how far the rounding-off drilling plans could be feasible. The second heuristic method fixes all capital expenditure decisions prior to production according to the solution of RMIP. The resulting restricted MEPPO problem can be decomposed into a series of capacitated drilling scheduling subproblems, each contingent on one leaf node of the composite scenario tree. All subproblems are guaranteed to be feasible and their solutions can be combined with the fixed capital decisions together to recover a feasible integer solution to the MEPPO. The third heuristics simply computes the minimal required capital costs and several other reward measures for each project and then uses them to form a few static Knapsack subproblems. The solution provides some insights to which projects should not be included in the portfolio.

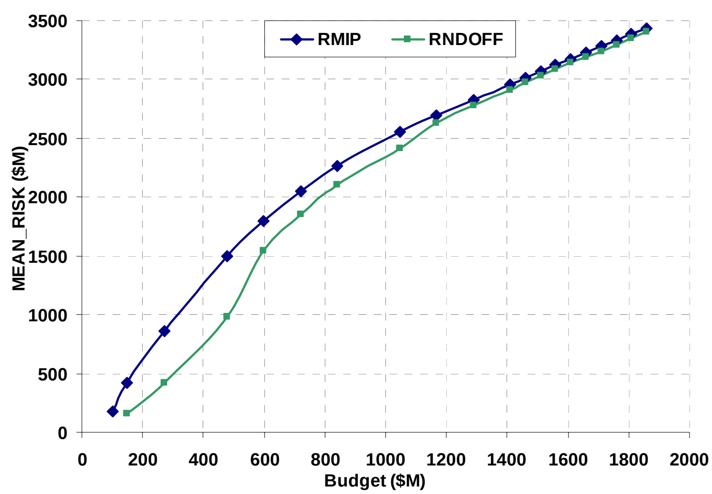
Our numerical experiments show that both the RNDOFF and the DCMP approaches work well in most cases. The RNDOFF will be used in a "zig-zag" procedure to globally search for solutions in a fast way and the DCMP will be used to find good integer solutions locally. Particularly, the gap between the RMIP and the DCMP is normally around 5%. For small or moderate instances, the gap often can be less than 1%. The NPSK subproblem can be used to accelerate both approaches when the initial budget is too tight to finance most combinations of those projects. We also found that the stringent budget cases are much more difficult to solve than those of sufficient budget.

Indirect solution: an inverse problem and a zig-zag bounding procedure

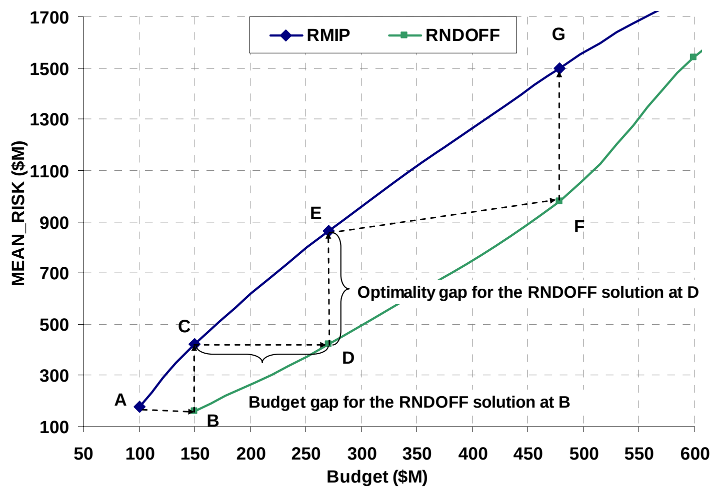
Once we have a rounding-off integer solution from the RMIP, we can see how far it can be feasible by solving an inverse problem, named RNDOFF. The RNDOFF problem are defined to have the same formulation as the MEPPO except that the initial budget is treated as a decision variable, all binary capital expenditure decisions and drilling decisions are fixed to the rounding-off solution of the RMIP, and everything else is kept unchanged. The RNDOFF problem turns out to be a simple problem since all decision variables but the inventory variables and the initial budget are fixed either explicitly or implicitly. Theoretically, the RNDOFF problem can be solved manually for small instances. Practically, we can simply just call GAMS/CPLEX to solve it within a few seconds. We call the optimal objective value of the RNDOFF problem as the RNDOFF minimal budget, which is the minimal required initial budget to finance the rounding-off drilling plans of the RMIP. Accordingly, we also call the round-off solution as the RNDOFF solution corresponding to the RNDOFF minimal budget. The RNDOFF minimal budget is greater or equal to the original (initial) budget.

Obviously, the RNDOFF solution is a feasible integer solution to the MEPPO problem when the initial budget is set to the RNDOFF minimal budget. The gap between the original budget of the MEPPO problem and the RNDOFF minimal budget provides a reasonable measure for how far away the RNDOFF solution to be an optimal integer solution of the original MEPPO problem. When the gap is zero or negligible, the RNDOFF solution can be treated as an optimal solution of the MEPPO problem. Otherwise, if the gap is sufficiently small such that the investor feels the budget increase is acceptable, the RNDOFF solution also provides a good reference. In this case, we would be interested in how well the RNDOFF solution for the MEPPO problem when the initial budget is set to the RNDOFF minimal budget. To answer this question, we can solve a new RMIP problem with the budget set to the RNDOFF minimal budget.

If we start at solving the RMIP with a small initial budget and repeat the above procedure until the budget level exceeds some high level, we can draw the two series of solutions of the RIMP and the RNDOFF as the following Figure 4-6 (a) and (b).



(a)



(b)

Figure 4-6 The solutions of the RMIP and the RNDOFF problem are generated in a "zig-zag" search procedure. The Figure 4-6 (a) shows the trend of the solutions in a wider budget region. The Figure 4-6 (b) elaborates the "zig-zag" search procedure.

Each point at the top curve in Figure 4-6 represents one solution of the RMIP for the corresponding initial budget. Since the solution relaxes the integral drilling decisions, it provides an upper bound for the original MEPPO problem with the same budget. For example, in Figure 4-6 (b), point A, C, E, and G are solutions of the RMIP problem and provide upper bounds for the MEPPO at corresponding budgets. All the points at the bottom curve in Figure 4-6, such as point B, D, and F, are integer solutions and feasible to the MEPPO. So the bottom curve provides a series lower bounds for the MEPPO. Consequently, the optimal integer solutions of the original problem MEPPO must be bounded somewhere between the two curves.

The two series of solutions of the RMIP problem and the RNDOFF problems can be generated by a fast "zig-zag" search procedure, which is described below. In the zig-zag procedure, we alternatively solve a series of RMIP and RNDOFF problems, one after the other. The solution of an RMIP problem is the start point for the succeeding RNDOFF problem. The solution of the RNDOFF will serve as the starting point for the next RMIP problem as well, and so far so forth. For a properly chosen starting point whose budget level is not so small that none of the projects will be taken, the budget level is monotonically increased during the zig-zag process.

Let's start with a small budget, e.g. the budget at point A. We relax the drilling decisions and solve the RMIP problem with this budget. The optimal solution is drawn as the point A, which provides an upper bound for the MEPPO problem at budget level A. We can round down the drilling decisions to get a rounding-off integer solution for the MEPPO. The rounding-off solution is most likely infeasible. We stick to this rounding-off solution and solve the corresponding RNDOFF problem as defined previously to find the minimal required budget that makes the rounding-off solution feasible to the MEPPO. Then, we solve the MEPPO once with all integer solutions fixed to the integer rounding-off solution and substitute the budget with the minimal RNDOFF budget. The restricted MEPPO can be solved very fast, usually taking a few seconds or so. The solution is drawn as the point B in Figure 4-6 (b). Since the rounding-off solution is integer feasible for the MEPPO problem at budget level B, it serves as a lower bound for the budget level B.

We repeat the above process by solving the RMIP at budget level B to get an upper bound at point C, and then fix the rounding-off solution at point C to find the minimal required budget at point D, and so on so forth. We will alternatively obtain the points A→B→C→D→E→F→G→… in sequence. Note that in Figure 4-6 (b), any two points on the same vertical transition line have the same budget and any two points on the same (sloped but roughly) horizontal transition line represent a solution with fractional drilling decisions and its rounding-off integer solution. Based on those points, we draw two curves in Figure 4-6 to show the trend of the sequence of upper bounds and the sequence of lower bounds as generated by the zig-zag procedure. Clearly, the optimal integer solution of the MEPPO must lie somewhere between the two curves.

Although the zig-zag procedure can not exactly solve the MEPPO problem, it does provide an intuitive and visual way to understand how the optimal solution would evolve as the initial budget varies. In order to solve the MEPPO directly for a pre-specified initial budget level, we provide an efficient decomposition method to look for good integer drilling plans, based on the relaxed solutions of the RMIP.

Direct solution: a decomposition method

Geological uncertainties of an E&P project are mainly involved in the early phases of its lifecycle. In our E&P model, we assume that geological uncertainties are resolved prior to the production phase. As a result, if the capital expenditure decisions, such as the acquisition, appraisal, and capacity installation, are determined, the remaining problem becomes separable scenario-based drilling scheduling problems. Each problem is independent to all others and specific to a scenario of the scenario tree.

Following the above observation, we propose a decomposition method (called DCMP) to recover an integer feasible solution for the MEPPO problem based on the RMIP solution. The RMIP solution is obtained by solving the MEPPO problem with relaxing the integral drilling decisions and keeping other binary decisions unchanged. In the RMIP solution, fractional drilling decisions are used to assist the solver to assess the productivity of the fields and determine the optimal capital investment decisions in a fast way. It is reasonable to think that the binary capital investment decisions of the RMIP solution are consistent estimates to those of the MEPPO solution. Based on this view, we propose a decomposition method as follows.

To facilitate the following discussion, we use a different way to divide the lifetime of an E&P project into two parts, the exploration part and the production part. The exploration part includes acquisition and all exploratory activities up to the completion of facility construction and prior to drilling the first production well. During this part, decisions are made with incrementally refined information about the geological uncertainties. Geological uncertainties are assumed to be "fully unfolded" prior to the production. From Figure 4-5, we can see that the first part decisions are non-anticipative. In other words, these decisions are made with partial information. In contrast, the decisions involved in the second part, or the production part, are made with complete information. So once all the first part decisions are given, the remaining part becomes a series of deterministic drilling and production planning problem.

In reality, the four tank model parameters will never be truly revealed until the last drop of oil is depleted. We assume that, up to the starting point of production, the thus much observed information is sufficient for us to forecast the field performance and estimate what future production decisions will likely be, given those information. We further assume additional information about the four parameters would become more and more marginal as for helping make better today's decisions so that we can simply treat them as deterministic after the start of production. After all, the purpose of learning is to make better acquisition decision. We can always add more phases for observations and learning. However, this will lead the problem to become much harder to solve.

We can design different heuristic methods to determine the first part decisions. One good choice as we would use is to fix the first part decisions according to the optimal solution of the RMIP problem. In such a way, we use the RMIP solution to help select oil fields and configure the initial production units for them. Once the first part decisions of the original MEPPO problem are fixed, all the sample paths of the underlying composite scenario tree are completely decoupled. The sample paths are determined by the leaf nodes of the composite scenario tree. So the restricted MEPPO problem whose the first part decisions are fixed can be decoupled into $|\mathbb{S}|\cdot|\mathbb{N}_3|$ independent subproblems, called drilling scheduling subproblems.

Each subproblem tries to schedule the drilling and production plans for selected oil fields, subject to a common budget constraint and separate production capacity constraints which are determined by the first part decisions. Although each subproblem only involves a small number of discrete variables and can be solved very quickly, the overheads to generate and solve $|\mathbb{S}|\cdot|\mathbb{N}_3|$ MIP problems are still too expensive to be efficient. Instead of decomposing the restricted MEPPO problem by decoupling all $|\mathbb{S}|\cdot|\mathbb{N}_3|$ leaf nodes of the composite scenario tree, we can only decouple either one of the two subtrees of the full composite scenario tree. Thus, we only need generate either $|\mathbb{S}|$ or $|\mathbb{N}_3|$ subproblems. Our numeric experiments have shown that the saved time by reducing the overheads dominates the solution time for moderate and large instances.

We choose to decouple the first subtree, i.e., the subtree corresponding to the combinatorial "Pay/No pay" states, and decompose the restricted MEPPO problem into $|\mathbb{S}|$ subproblems, called DCMP subproblems. Therefore, after we have solved the RMIP problem, we can recover an integer feasible solution for the original MEPPO problem by solving $|\mathbb{S}|$ independent subproblems, one by one, defined as follows.

For consistence, we use the same notations of decision variables involved in the DCMP subproblem as their counterparties in the MEPPO problem. To differentiate them, we drop the superscript s of the decision variables for the DCMP subproblem. All those variables and constraints of the subproblem are involved in the production phase between time stage t_3 and the final stage $T = t_5$ of the MEPPO problem.

Given an RMIP solution, for each state $s \in \mathbb{S}$, we define and initialize a few parameters as below,

$$\widehat{M}_i^{n_3} = M_i^{s,n_3} \cdot \sum_{l \in L} x\_cap1_{i,l}^{a(s,n_3)}, \qquad \forall n_3 \in \mathbb{N}_3, \tag{4.4.1}$$

$$U_i^{n_3} = y\_cap1_i^{a(s,n_3)}, \qquad \forall n_3 \in \mathbb{N}_3, \tag{4.4.2}$$

$$\widehat{B}_{t_3}^{n_3} = \left(1 + r_0\right) \cdot y\_inv_{t_3-1}^{a_{t_3-1}(s,n_3)}, \qquad \forall n_3 \in \mathbb{N}_3, \tag{4.4.3}$$

in which $x\_cap1_{i,l}^{a(s,n_3)}$, $y\_cap1_i^{a(s,n_3)}$, and $y\_inv_{t_3-1}^{a_{t_3-1}(s,n_3)}$ are extracted from the RMIP solutions. The three collections of parameters define the state-n_3 specific maximal allowed well numbers, initial production capacity, and initial budget which for the state s-dependent DCMP subproblem.

(DCMP subproblem)

$$z^{sub}\left(\rho\right) = \text{Max} \quad y\_expval - \rho \cdot y\_expdev, \tag{4.4.4}$$

s.t.

$$\sum_{t=t_3}^{T-1} x\_drill_{i,t}^{n_3} \le \widehat{M}_i^{n_3}, \qquad \forall i, n_3, t_3 \le t < T \tag{4.4.5}$$

$$y\_capex_{i,t}^{n_3} = DrillCost_i \cdot x\_drill_{i,t}^{n_3}, \qquad \forall i, n_3, t_3 \le t < T,\ t \ne t_4 \tag{4.4.6}$$

$$y\_prod_{i,\tau,t}^{n_3} = Q_{i,t-\tau}^{n_3} \cdot x\_drill_{i,\tau}^{n_3}, \qquad \forall i, n_3, t_3 \le \tau \le t \le T \tag{4.4.7}$$

$$y\_total_{i,t}^{n_3} = \sum_{\tau=t_3}^{t} y\_prod_{i,\tau,t}^{n_3}, \qquad \forall i, n_3, t_3 \le t \le T \tag{4.4.8}$$

$$y\_opex_{i,t}^{n_3} = ProdCost_i \cdot y\_total_{i,t}^{n_3}, \qquad \forall i, n_3, t_3 \le t \le T \tag{4.4.9}$$

$$\sum_{k \in K} x\_cap2_{i,k}^{n_3} \le 1, \qquad \forall i, n_3 \tag{4.4.10}$$

$$y\_capex_{i,t_4}^{n_3} = \sum_{k \in K}\left(CapCost2_k \cdot x\_cap2_{i,k}^{n_3}\right) + DrillCost_i \cdot x\_drill_{i,t_4}^{n_3}, \qquad \forall i, n_3 \tag{4.4.11}$$

$$y\_cap2_i^{n_3} = \sum_{k \in K} Cap2_k \cdot x\_cap2_{i,k}^{n_3}, \qquad \forall i, n_3 \tag{4.4.12}$$

$$y\_total_{i,t}^{n_3} \le \widehat{U}_i^{n_3}, \qquad \forall i, n_3, t_3 \le t \le t_4 \tag{4.4.13}$$

$$y\_total_{i,t}^{n_3} \le \widehat{U}_i^{n_3} + y\_cap2_i^{n_3}, \qquad \forall i, n_3, t_4 < t \le T \qquad (4.4.14)$$

$$y\_rev_{i,t}^{n_3} = F_{0,t} \cdot y\_total_{i,t-1}^{n_3}, \qquad \forall i, n_3, t_3 + 1 \le t \le T-1 \quad (4.4.15)$$

$$y\_rev_{i,T}^{n_3} = F_{0,T} \cdot \left(y\_total_{i,T-1}^{n_3} + y\_total_{i,T}^{n_3} \right), \quad \forall i, n_3 \qquad (4.4.16)$$

$$\sum\nolimits_i y\_capex_{i,t_3}^{n_3} + (1 - r_{tax}) \cdot \sum\nolimits_i y\_opex_{i,t_3}^{n_3} + y\_inv_{t_3}^{n_3} = \widehat{B}_{t_3}^{n_3}, \forall n_3 \qquad (4.4.17)$$

$$\begin{aligned} &\sum\nolimits_i y\_capex_{i,t}^{n_3} + (1 - r_{tax}) \cdot \sum\nolimits_i y\_opex_{i,t}^{n_3} + y\_inv_t^{n_3} = \\ &\qquad (1 + r_0) \cdot y\_inv_{t-1}^{n_3} + \left(1 - r_{royal}\right) \cdot (1 - r_{tax}) \cdot \sum\nolimits_i y\_rev_{i,t}^{n_3}, \\ &\qquad\qquad \forall n_3, t_3 \le t \le T-1 \end{aligned} \qquad (4.4.18)$$

$$\begin{aligned} (1 - r_{tax}) \cdot \sum\nolimits_i y\_opex_{i,T}^{n_3} + y\_inv_T^{n_3} &= (1 + r_0) \cdot y\_inv_{T-1}^{n_3} \\ &+ \left(1 - r_{royal}\right) \cdot (1 - r_{tax}) \cdot \sum\nolimits_i y\_rev_{i,T}^{n_3}, \ \forall n_3 \end{aligned} \qquad (4.4.19)$$

$$y\_expval = \sum\nolimits_{n_3} Pr1^{n_3} \cdot y\_inv_T^{n_3}, \qquad (4.4.20)$$

$$y\_inv_T^{n_3} - y\_expval - u^{n_3} + v^{n_3} = 0, \qquad \forall n_3 \qquad (4.4.21)$$

$$y\_expdev = \sum\nolimits_{n_3} Pr1^{n_3} \cdot v^{n_3}. \qquad (4.4.22)$$

We have ignored the variable definitions. The relevant information can be conveniently recovered from the MEPPO problem without confusion. For example, the decision variables' types of the DCMP subproblem are the same as the MEPPO problem.

The correspondences between the constraints of the DCMP subproblem and those of the MEPPO are as follows. The objective function (4.4.4) corresponds to (4.3.24), (4.4.5) ~ (4.4.16) to (4.3.7) ~ (4.3.18), (4.4.17)~(4.4.19) to (4.3.21) and (4.2.22), and (4.4.20) ~ (4.4.22) to (4.3.25) ~ (4.3.27), respectively.

Assume we have solved the RMIP problem. If the solution is non-trivial, i.e., all decisions being zero, we can find an integer feasible solution to the MEPPO as follows.

For each first subtree state $s \in \mathbb{S}$, we perform the following three actions:

1. Initialize the three collections of parameters according to the solution of RMIP for the state *s*.

2. Solve the DCMP subproblem (4.4.4) ~ (4.4.22) to find an integral solution $\left\{x\_drill_{i,t}^{n_3}, \forall i, n_3, t_3 \le t < T\right\}$ and $\left\{x\_cap2_{i,k}^{n_3}, \forall i, n_3, k\right\}$. If the solver doesn't return a solution of the DCMP subproblem within the designated execution time, we set the two collections of integral variables to zero.

3. Fix the state-*s*-contingent components of the remaining integral variables of the MEPPO problem by

$$\begin{aligned} &\left\{x\_drill_{i,t}^{s,n_3}, \forall i, n_3, t_3 \le t < T\right\}^{MEPPO} \leftarrow \left\{x\_drill_{i,t}^{n_3}, \forall i, n_3, t_3 \le t < T\right\}^{DCMP\ subproblem} \\ &\left\{x\_cap2_{i,k}^{s,n_3}, \forall i, n_3, k\right\}^{MEPPO} \leftarrow \left\{x\_cap2_{i,k}^{n_3}, \forall i, n_3, k\right\}^{DCMP\ subproblem}. \end{aligned}$$

4. Return to 1 for the next element $s \in \mathbb{S}$ until all elements in $\mathbb{S}$ have been visited.

After the return of the above loop, all integral decision variables of the MEPPO should have been fixed to proper values. Those values are guaranteed to be feasible unless the RMIP hasn't returned a feasible solution at the start. The resulting restricted MEPPO problem becomes a pure linear programming problem. Particularly, many continuous variables are implied to be fixed as well. So the restricted MEPPO can be solved very quickly. Indeed, the restricted MEPPO usually can be solved during the pre-process procedure by eliminations.

Our computational experiments have shown that the integer DCMP solutions are usually good and stable enough to be near-optimal solutions to the MEPPO problem. The gap between the DCMP solution and the RMIP solution is usually around 5% for most cases and up to 20% for difficult cases in which tight initial budgets are met.

Pre-screening heuristics: further scenario reduction and Knapsack heuristics

The MEPPO problem is particularly difficult to solve when the budget is too tight to simultaneously finance the exploration of the majority of all projects. In this case, even the RMIP may not return a meaningful solution after tens of minutes of execution. As we have observed that if the RMIP can be solved efficiently at least we can find some good integer solution based on the previous DCMP method. As a consequence, we need seek other methods to speed up the RMIP solution process, usually at the cost of sacrificing numerical accuracy to some extent.

As we mentioned before, the number of projects is one major factor affecting the complexity of the MEPPO problem. If we can identify one or more "poor" projects according to certain rule(s), we can exclude them and solve a reduced instance of the MEPPO problem. We proposed two heuristic approaches to achieve this goal.

The first idea is to (further) reduce the composite scenario tree to such a small scale that we can solve the RMIP problem in seconds or tens of seconds. We can try a few times with different reduction configurations to identify which projects are significantly more likely to be excluded. We disable those projects and solve the reduced MEPPO problem with previous methods.

The second idea is to use a pair of proper cost and reward measures to characterize individual E&P projects. For example, one cost measure could be the ENPV of the minimal required capital costs for a field to start production. Since a field starts production only if it has been acquired, appraised, and at least equipped with a platform of the minimal capacity, all of those costs together impose the minimal required capital for the field to produce. We can use the expected recoverable reserve or the ratio of the expected per well recoverable reserve to the expected per well production cost. Other reasonable measures are possible and encouraged. We then run the static capital budgeting problem for different choices of the cost/reward measures. We will maintain a list which ranks projects according to the likelihood to be exclusion.

4.4.2 Methods to verify the model implementation

The MEPPO model has a complex structure involving multistage decisions and disparate uncertainties. Even moderate instances lead to hundreds of thousands of decision variables and the similar magnitude of constraints. After we have implemented the model in mathematical programming languages like GAMS and finished debugging, the first question comes to our mind is how to ensure the implementation is correct and the resulting solution does solve the original MEPPO model. We propose a few simple methods to verify the correctness of the model and solutions.

Simple feasibility check

The MEPPO model has a property that it always has a feasible solution which is to abandon all projects from the beginning and not to make any investment decisions. So if we mandatorily set all acquisition decisions to be zero, the model must be feasible and the portfolio ENPV must equal the initial budget and no cost or revenue term should be involved along any sample path. Otherwise, one of the logic consistence constraints must be incorrectly modeled.

Deterministic MEPPO model check

Our MEPPO implementation is a quite flexible and allows users to choose the fanning factors to generate a user-defined scenario tree. We can mandate the scenario tree to degenerate to be a single sample path by forcing all fanning factors to be one. In such a way, our stochastic MEPPO model degenerates to a deterministic dynamic capital budgeting model, whose solution and optimality is much easier to verify.

Individual decision tree solution check

The MEPPO model can degenerate to a single project model for a specific oil field by disabling other projects' acquisition decisions and using a sufficiently large initial budget. The resulting single project MEPPO model is essentially a single decision tree model, which can be solved very efficiently. It is easier to check the solution with a single project than to check the solution with multiple projects. There are different ways to check the single decision tree solutions. The approach we use is to build a decision tree model with the well-known decision tree modeling tool, DPL, according to our phased E&P model and chosen parameters. We can solve the decision tree in both systems. We can perform this comparison for each individual oil field, one by one. If the solution structures of those two approaches coincide, the MEPPO model should be correct since the DPL model is relatively easier to check.

Spreadsheet based sample path solution check

The above methods are effective and convenient. However, they only provide assurance for the MEPPO model under some simple scenarios. The following sample-path based verification method allows us to check the MEPPO model under actual working status without making any sacrifice in model completeness or solution quality.

After we have solved the MEPPO model and found a feasible integer solution, we can extract all capital investment decisions and operating decisions along any specified sample path. According to those solutions and model inputs, we can re-calculate all incurred cash flows according to the MEPPO mathematical model. From the re-calculated cash flows, we can recover inventory cash positions at each time point along the path. Any negative inventory position indicates something must be wrong. Otherwise, we can continue the following verification procedure. We compare the re-calculated cash flows with the cash flows from the optimal solution along the same sample path. If they coincide, the model passes the verification test for the specified sample path. Otherwise, a mismatch is found and reported. The verification fails.

We can repeat the above verification procedure for a number of randomly selected sample paths. Although in theory we can say the model is absolutely correct only if we have tried all possible inputs and verified each solutions along all sample path, in practice, a small number of sample-path tests already enough to tell the correctness of the implementation with sufficiently high credibility. The reason is that most investment decisions are discrete and so do their incurred cash flows. Those cash flows are involved in the exact equalities of budget constraints. Any mistake of the model would lead to apparent mismatches in more than one location.

4.4.3 Numerical experiments

We implement the MEPPO model in GAMS. The complete MEPPO model consists of three modules, a scenario generation (SCENGEN) module, an optimization model module, and a solution reporting module.

The SCENGEN module allow us to reconfigure the MEPPO model. The configurations include the number of oil fields, the ranges of the model parameters, the decision stages, and the scenario tree structure. The module further permits us to define which parameter(s) are observable at a decision stage and can be used to update beliefs on other parameters of the same stage. New parameters can be added conveniently.

The two foundations of the scenario generator are Monte Carlo simulation and scenario reduction. We will introduce the details of the scenario generator in Chapter 5.

The optimization model module includes the model implementation, solution algorithms, and some model verification codes. The module also includes iterative codes to study the impact of the different budget levels to the optimal portfolio strategies.

The reporting module collects useful solution information, organizes them in a structured way and sends them to spreadsheets of a Microsoft Excel template file.

Test problem: a five-project portfolio model

Our MEPPO model and solution methods can deal with up to ten projects within an acceptable running time frame (less than twenty minutes for most cases). For simplicity, we use a five-project portfolio to perform extensive numerical experiments.

The numerical tests are based on five randomly generated projects. The following Table 4-4 shows two marginal statistics (mean and standard deviation) for the parameters of the five oil fields. The parameters include the field recoverable reserve, reserve per well, well initial production rate, well decay rate, and chance of success. The former four parameters are associated with the individual tank models.

Table 4-4 Statistics of the tank model parameters

Parameter \ Field		i1	i2	i3	i4	i5
Reservoir reserve size	mean	361.9805	40.7086	208.3981	146.9992	571.3176
(MMbbl)	stdev	1085.9413	122.1258	625.1943	440.9976	1713.9526
Well reserve size	mean	1.8545	3.2833	5.7040	2.3533	6.7354
(MMbbl)	stdev	3.7090	6.5666	11.4081	4.7067	13.4707
Init. production rate	mean	0.8698	1.5249	1.4605	1.0088	0.7595
(bbl/year)	stdev	1.7397	3.0499	2.9210	2.0177	1.5191
Decay rate	mean	0.4690	0.4644	0.2560	0.4287	0.1128
(1/year)	stdev	0.9381	0.9289	0.5121	0.8574	0.2255
Chance of success	(%)	61.4%	17.7%	44.1%	35.3%	25.1%

The parameters of the tank model are usually assumed to follow multi-lognormal distributions. In other words, the logged parameters follow multinormal distributions. The Table 4-5 gives the corresponding marginal statistics of the logged parameters.

Table 4-5 Statistics of the tank model parameters (logged data)

Logged Parameter \ Field		i1	i2	i3	i4	i5
Reservoir reserve size	mean	4.7403	2.5551	4.1882	3.8391	5.1967
(MMbl)	stdev	1.5174	1.5174	1.5174	1.5174	1.5174
Well reserve size	mean	-0.1871	0.3841	0.9365	0.0511	1.1027
(MMbl)	stdev	1.2686	1.2686	1.2686	1.2686	1.2686
Init. production rate	mean	-0.9442	-0.3828	-0.4259	-0.7959	-1.0798
(bbl/year)	stdev	1.2686	1.2686	1.2686	1.2686	1.2686
Decay rate	mean	-1.5618	-1.5716	-2.1671	-1.6517	-2.9872
(1/year)	stdev	1.2686	1.2686	1.2686	1.2686	1.2686

The values in the above two tables are consistent to each other. The equivalent relationships are guaranteed by the following propositions.

Proposition 4.1 Let normal random variable $X \sim N\left(\mu,\sigma^2\right)$, then $\mathbb{E}\left[X\right]=\mu$ and $\mathbb{VAR}\left[X\right]=\mathbb{E}\left[X^2\right]-\left(\mathbb{E}\left[X\right]\right)^2=\sigma^2$.

Proposition 4.2 Let lognormal random variable $Y \sim LN\left(\mu,\sigma^2\right)$, then Y>0, $\log\left(Y\right)\sim N\left(\mu,\sigma^2\right)$, $\mathbb{E}\left[\log\left(Y\right)\right]=\mu$, and $\mathbb{VAR}\left[\log\left(Y\right)\right]=\sigma^2$.

From the above two propositions, we have the following Proposition 4.3.

Proposition 4.3 Let lognormal random variable $Y \sim LN\left(\mu,\sigma^2\right)$, Y>0, $\mathbb{E}\left[Y\right]=\mu_Y$, and $\mathbb{VAR}\left[Y\right]={\sigma_Y}^2$, then we have

1) $\mu_Y = e^{\mu+\frac{1}{2}\sigma^2}$ and $\sigma_Y{}^2 = e^{2\mu+\sigma^2}\left(e^{\sigma^2}-1\right)$,

2) $\mu = \log\left(\mu_Y\right) - \frac{1}{2}\log\left(\frac{\sigma_Y^2+\mu_Y^2}{\mu_Y^2}\right)$ and $\sigma^2 = \log\left(\frac{\sigma_Y^2+\mu_Y^2}{\mu_Y^2}\right)$.

The propositions 4.1 through 4.3 are well-known results. The proofs are straightforward and can be found on most statistics textbooks. Those results are useful to calculate marginal statistics from either direction. A similar result in closed-form for correlated multi-lognormal random vectors can be found in the appendix of Wang (1998).

The following Table 4-6 through Table 4-8 list relevant cost parameters.

Table 4-6 Field E&P cost parameters

Parameters \ Field		i1	i2	i3	i4	i5
Acquisition cost	(MM$)	7.00	6.40	16.10	14.70	4.00
Appraisal cost	(MM$)	60.60	39.10	21.00	39.30	30.20
Drilling cost	(MM$/well)	10.70	10.40	12.70	21.60	11.20
Production cost	($/bbl)	17.90	29.50	29.80	15.70	29.00

Table 4-7 Optional platforms with construction cost and initial capacity

		small	mid	large
Installed capacity	(MMbbl)	25	50	75
Installation cost	(MM$)	15	35	50

Table 4-8 Optional expansion plans with expansion cost and expanded capacity

		small	mid	large
Expanded capacity	(MMbbl)	20	40	60
Expansion cost	(MM$)	15	32	50

Numerical experiments

We perform numerical experiments under three controlled factors, the scenario tree size, the initial budget level, and the risk factor (ρ). The control over the composite scenario tree is realized by performing an optional scenario reduction to the second subtree (Scen2) to ensure the resulting problem is solvable. Without this scenario reduction, even the relaxed RMIP may not be practically soled within an acceptable time.

Table 4-9 lists statistics of some MEPPO instances when the terminal nodes of the second scenario tree are reduced to 30, 50, 75, and 100. The table gives the number of the composite scenario tree (in column 3), the number of various decision variables (in column 4 ~7), the number of constraints, and the number of nonzero elements.

Table 4-9 Example sizes under different second scenario trees

Scenario Tree Size			Variable numbers				Constraint number	Nonzero elements
Scen1	Scen2	Terminal	Binary	Integer	Continuous	Total		
13	30	390	295	5,220	61,375	66,890	63,131	190,275
13	50	650	295	8,700	102,068	111,063	104,782	315,020
13	75	975	295	13,050	152,046	165,391	156,233	469,302
13	100	1,300	295	17,300	202,124	219,719	207,684	623,300

From the Table 4-9, we can see that the majority of discrete variables come from the integral drilling decisions. This observation motivates us to relax them to develop the RMIP based heuristic methods (RNDOFF and DCMP). From the last two columns, we know on average each constraint roughly involves three decision variables (three nonzero elements). This is because many constraints are used to model proper logic and timing relationships among two or more decisions. Such constraints can be efficiently substituted with implied bounds to variables by the preprocess procedure of LP solvers.

The following three tables, Table 4-10, Table 4-11, and Table 4-12, collect the solution and solution times for various combinations of the three factors. We can see from Table 4-10 that the solutions for different Scen2 sizes are consistent to each other. Particularly, the differences become much slimmer when the budget rises high enough, for each select risk factor ρ.

Table 4-10 The optimal objective value of the integer DCMP solutions (Unit: MM$)

Scen2	30			50			75			100		
Budget (MM$) \ ρ	0	0.2	0.4	0	0.2	0.4	0	0.2	0.4	0	0.2	0.4
100	194.7	177.4	160.9	171.6	155.8	145.1	87.3	85.4	79.8	161.0	83.2	71.9
150	363.3	336.5	312.0	350.3	325.2	267.9	304.7	326.8	301.7	271.3	291.5	269.8
200	489.0	459.9	430.9	467.5	438.4	408.2	434.4	442.0	378.5	422.7	444.4	414.0
250	601.3	570.9	534.0	582.3	546.2	517.7	490.0	560.1	522.4	604.3	519.1	495.1
300	710.9	673.9	639.8	683.0	634.9	595.5	686.3	656.2	606.3	615.9	662.2	614.8
350	812.4	776.8	737.2	771.3	727.9	687.0	798.2	751.9	706.1	659.1	737.0	704.0
400	908.7	860.7	819.6	864.0	815.1	769.1	882.5	840.2	790.7	899.3	846.6	778.2
450	993.7	948.5	900.4	945.1	896.5	846.2	972.4	918.7	867.3	981.1	928.4	873.1
500	1072.1	1026.7	978.4	1028.1	976.2	925.6	1039.0	999.4	914.7	1063.9	980.1	948.2
550	1143.0	1094.6	1046.6	1097.4	1046.1	992.5	1128.4	1070.6	1011.8	1139.3	1079.6	1018.3
600	1213.2	1163.4	1107.1	1167.7	1112.7	1054.9	1194.7	1118.7	1076.9	1205.7	1146.6	1084.1
650	1277.5	1226.4	1174.9	1236.1	1177.8	1120.7	1263.2	1201.0	1139.4	1277.0	1213.2	1148.3
700	1337.3	1284.4	1231.1	1296.7	1237.9	1181.4	1325.8	1261.0	1198.2	1338.1	1274.8	1208.4
750	1405.9	1351.7	1296.6	1362.9	1303.1	1243.0	1389.1	1324.1	1259.2	1401.6	1338.0	1269.6
800	1472.6	1416.4	1360.2	1426.4	1365.5	1304.5	1451.5	1385.6	1319.5	1467.4	1398.3	1328.7
850	1540.2	1483.2	1425.2	1488.3	1428.0	1366.2	1514.4	1447.3	1379.8	1525.7	1452.9	1389.0
900	1603.4	1544.8	1485.7	1550.6	1487.6	1423.5	1575.7	1507.1	1438.2	1591.4	1519.4	1447.6
950	1668.8	1608.7	1548.6	1612.8	1547.4	1484.7	1635.9	1565.8	1496.1	1651.3	1578.8	1505.9
1000	1728.8	1668.4	1608.2	1670.5	1606.0	1541.3	1694.3	1624.9	1553.2	1709.1	1636.2	1563.3
1050	1790.1	1728.6	1666.7	1730.8	1665.4	1600.0	1753.4	1680.4	1610.2	1766.4	1692.7	1619.0

The DCMP solution procedure consists of two steps. The first step solves the RMIP problem and the second step solves a series of decoupled DCMP subproblems, each of which is a scenario-specific and deterministic drilling scheduling problem. Table 4-11 and Table 4-12 list the solution times for the RMIP problem and the DCMP subproblems. Normally, we choose a proper (relative) optimality tolerance[6] to allow

[6] In our case, we use 5% as the relative optimality tolerance. In GAMS, this is achieved by setting the option optcr to 5%.

the solver to return to avoid wasting time on the tailing effect. However, for some range of parameters, even reaching the optimality tolerance requires an unexpectedly long time, e.g., the RMIP problem with a very limited budget, 100 MM$ and 150 MM$ in Table 4-11. In such cases, we mandate the solver to prematurely return with best available solution by applying a running time budget to each execution. We have set the maximal allowed run time to be 20 minutes, or 1200 seconds. We can see that all instances in Table 4-11 with a budget less than 200 MM$ will be forcefully exited before the optimality tolerance reaches 5%. The actual optimality tolerances of the RMIP problem can be found in Tablee 4-12.

From Table 4-12 and Table 4-13, we can see that the solution of the RMIP problem is the bottleneck for the DCMP algorithm. Once the RMIP is solved, the DCMP subproblems can be solved efficiently. Particularly, the solution performance of the DCMP subproblems is relatively stable and irrelevant to the budget level and the risk factor. It is mainly affected by the scenario tree size and roughly in proportion.

Our preceding discussions suggests that in the future, in order to solve larger instances of the MEPPO problem, we need find more efficient ways to deal with the RMIP problem. This is beyond this dissertation and could be a future research topic.

Table 4-11 The solution time of the RMIP problem (Unit: seconds)

Scen2	30			50			75			100		
Budget (MM$) \ ρ	0	0.2	0.4	0	0.2	0.4	0	0.2	0.4	0	0.2	0.4
100	1201	1201	1201	1201	1203	1202	1202	1204	1204	1203	1206	1206
150	1201	1201	1201	1201	1202	1202	1203	1204	1205	1203	1208	1208
200	41	1138	1201	82	145	131	127	323	1205	307	547	735
250	25	64	88	60	114	144	150	311	445	287	491	531
300	23	58	56	66	147	144	99	366	320	275	422	686
350	19	46	53	77	183	144	132	251	387	221	402	424
400	16	41	39	66	198	115	123	202	247	215	376	453
450	13	29	34	66	169	106	117	162	365	225	286	310
500	17	28	25	92	145	90	98	152	359	192	305	373
550	12	24	23	70	109	102	94	148	328	199	302	330
600	11	27	29	58	71	94	116	179	322	185	326	324
650	11	22	25	76	57	71	110	196	330	202	412	371
700	13	23	23	60	56	77	110	154	359	177	288	297
750	9	17	16	47	49	65	126	129	288	185	371	339
800	12	22	23	48	52	86	112	150	249	213	340	325
850	11	18	22	52	54	73	104	167	182	239	232	305
900	13	22	22	48	92	49	113	243	188	198	289	293
950	9	17	25	42	59	57	93	120	255	208	262	239
1000	12	15	18	51	69	53	116	133	191	213	244	256
1050	10	16	17	44	71	57	106	119	200	235	266	295

Table 4-12 The solution time of the DCMP subproblems (Unit: seconds)

Scen2	30			50			75			100		
Budget (MM$) \ ρ	0	0.2	0.4	0	0.2	0.4	0	0.2	0.4	0	0.2	0.4
100	8	10	9	13	21	16	14	42	39	148	105	82
150	12	14	22	16	22	30	88	33	45	286	81	76
200	8	10	10	16	22	19	96	51	72	150	93	85
250	9	13	11	16	19	16	140	62	67	30	253	280
300	9	12	12	16	27	23	28	58	50	159	88	100
350	9	11	10	21	32	23	39	86	100	230	209	207
400	10	12	12	21	36	32	145	50	87	51	80	176
450	9	11	11	20	28	22	31	72	115	42	89	93
500	9	11	9	23	21	18	146	129	149	41	180	77
550	8	20	9	22	21	20	27	41	89	34	67	61
600	8	10	10	25	19	18	25	106	81	43	116	69
650	8	10	23	24	21	20	31	37	74	36	80	58
700	12	10	9	24	16	16	22	32	72	34	68	64
750	11	10	9	17	15	17	23	25	57	26	51	54
800	9	10	9	18	15	15	22	41	44	32	47	47
850	10	10	10	16	36	20	29	43	35	22	43	46
900	11	9	9	16	17	15	21	28	48	28	32	29
950	13	9	9	14	15	15	22	40	44	25	42	43
1000	11	9	8	16	16	14	20	20	34	25	33	29
1050	13	9	9	15	15	15	20	20	31	27	26	29

Solution quality

The MEPPO problem is solved as a large scale deterministic MIP. Like most large MIP problems, it is difficult to find and verify an integer optimal solution for the MEPPO problem. Usually we will be satisfied with a near-optimal solution which is integer feasible and has a sufficiently small optimality gap. The optimality gap measures the quality of the current best integer feasible solution (BF). It relies on an upper bound on the optimal value of the MIP. Normally the upper bound is estimated by a tight dual feasible solution or an optimal solution of some good relaxation problem. The upper bound indicates how good the best possible integer solution (BP) can be.

Different solvers[7] may use slightly different definitions to calculate the (relative) optimality gap. Our choice of the optimality gap calculation for the current best integer feasible solution (BF) is given by

$$Gap = \frac{|BP| - |BF|}{|BP|}. \tag{4.4.23}$$

In GAMS, after the return of a successful MIP solver call, the model attribute *Objest* provides an upper bound estimate of the best possible integer solution. We can use it to compute the relative optimality gap of the RMIP solution by

$$RMIP\_gap = \frac{|RMIP\_Objest| - |RMIP\_BF|}{|RMIP\_Objest|}, \tag{4.4.24}$$

[7] CPLEX calculates the relative optimality gap by appending a small strictly positive constant +1.0E-10 to the denominator in (4.4.23).

where $|RMIP\_BF|$ denotes the objective function value of the current best integer solution of the RMIP problem.

In general, the MIP solver will stop trying to improve the current best integer solution and return the solution once the gap is smaller than the preset relative optimality tolerance (Optcr), i.e. $Gap < Optcr$.

Similarly, we can compute the relative optimality gap between the DCMP solutions and the best possible integer solution of the MEPPO problem. In our DCMP method, a natural upper bound for the DCMP solution is provided by solving the relaxed MEPPO problem, i.e., the RMIP problem. However, since the RMIP problem itself is a large scale MIP, the RMIP solution also comes with an optimality gap as shown in Table 4-13, which means the RMIP solution may not provide an effective upper bound for the MEPPO problem. We know that only the exact optimal solution of the relaxed problem ensures an upper bound for its corresponding MIP. If the RMIP solution gap is small enough, we can use the optimal value of the RMIP solution as an estimated upper bound for the best possible integer solution of MEPPO problem. Otherwise, we should use the RMIP problem's upper bound to compute the DCMP solution gap, i.e., the $|RMIP\_Objest|$ in (4.4.24). So the DCMP gap is defined by

$$DCMP\_gap = \begin{cases} \dfrac{|RMIP| - |DCMP|}{|RMIP|} & , RMIP\_gap \le 5\% \\ \dfrac{|RMIP\_Objest| - |DCMP|}{|RMIP\_Objest|} & , otherwise. \end{cases} \tag{4.4.25}$$

The following two tables, Table 4-13 and Table 4-14, show the corresponding optimality gaps for the RMIP solutions and the ultimate DCMP solutions, respectively.

Table 4-13 Optimality gap of the RMIP solution (Unit: %)

Scen2	30			50			75			100		
Budget (MM$) \ ρ	0	0.2	0.4	0	0.2	0.4	0	0.2	0.4	0	0.2	0.4
100	11.01	14.75	17.51	14.40	18.00	19.42	42.54	41.25	43.53	19.79	43.62	46.46
150	5.17	6.37	6.82	4.94	5.95	13.94	12.35	6.70	6.83	11.54	12.96	13.31
200	4.71	4.74	5.02	4.36	4.54	4.35	4.58	4.57	10.58	4.64	4.55	4.36
250	3.47	3.57	4.60	2.87	3.34	2.11	2.29	2.05	1.83	1.57	2.07	1.63
300	1.31	2.16	2.00	2.20	3.03	2.66	3.18	1.83	3.63	2.15	2.29	2.02
350	1.35	1.02	1.11	1.96	2.08	1.68	0.88	0.88	1.01	1.16	1.04	1.09
400	0.70	1.10	0.92	1.23	1.09	1.28	0.70	0.58	0.57	0.69	0.77	0.75
450	1.15	1.18	1.47	1.32	1.36	1.57	0.87	0.89	0.77	0.94	0.72	0.74
500	0.46	0.35	0.32	0.57	0.56	0.51	0.40	0.31	0.35	0.40	0.36	0.38
550	0.39	0.54	0.38	0.76	0.44	0.45	0.31	0.28	0.35	0.32	0.32	0.32
600	0.29	0.24	0.38	0.26	0.25	0.44	0.29	0.25	0.38	0.53	0.26	0.27
650	0.27	0.35	0.31	0.25	0.24	0.24	0.28	0.27	0.25	0.29	0.26	0.26
700	0.25	0.21	0.18	0.36	0.25	0.36	0.31	0.30	0.32	0.35	0.30	0.30
750	0.17	0.30	0.30	0.22	0.18	0.19	0.22	0.25	0.23	0.35	0.24	0.22
800	0.20	0.20	0.18	0.18	0.17	0.17	0.21	0.19	0.19	0.23	0.22	0.22
850	0.15	0.11	0.12	0.22	0.17	0.18	0.22	0.18	0.18	0.27	0.64	0.20
900	0.19	0.16	0.20	0.23	0.23	0.42	0.23	0.21	0.20	0.21	0.21	0.20
950	0.13	0.15	0.16	0.17	0.27	0.15	0.21	0.23	0.23	0.21	0.20	0.18
1000	0.14	0.13	0.13	0.22	0.23	0.20	0.22	0.15	0.20	0.24	0.20	0.17
1050	0.10	0.10	0.14	0.17	0.16	0.15	0.19	0.27	0.20	0.33	0.26	0.21

Table 4-14 Optimality gap of the DCMP solution (Unit: %)

Scen2	30			50			75			100		
Budget (MM$) \ ρ	0	0.2	0.4	0	0.2	0.4	0	0.2	0.4	0	0.2	0.4
100	14.04	12.56	12.28	15.91	15.19	14.23	46.86	43.99	42.01	20.81	44.08	45.78
150	6.20	6.06	6.16	4.50	4.47	7.31	12.46	4.70	4.81	25.60	9.72	9.26
200	5.02	5.02	5.29	5.40	5.17	5.78	13.48	5.03	6.83	17.29	5.44	5.31
250	5.59	5.96	6.21	4.88	4.78	5.12	23.55	4.36	4.55	4.52	12.71	10.88
300	6.35	6.00	6.13	4.67	5.32	5.98	4.31	3.98	3.63	17.29	3.65	4.46
350	5.65	5.59	5.74	4.96	5.05	5.25	3.99	4.20	4.27	23.77	6.78	4.82
400	5.34	5.76	5.92	4.44	4.94	4.94	4.52	3.88	3.93	3.64	3.67	5.89
450	4.80	4.88	5.03	4.10	4.25	4.39	3.47	3.59	3.61	3.45	3.37	3.38
500	5.56	5.52	5.69	4.14	4.26	4.32	4.88	3.45	6.58	3.32	5.94	3.32
550	6.13	5.91	5.98	4.31	4.60	4.66	3.52	3.54	3.58	3.24	3.14	3.19
600	6.37	6.31	6.70	4.83	4.84	5.05	3.98	5.41	3.65	3.43	3.33	3.24
650	6.53	6.37	6.32	4.70	4.84	4.86	3.83	3.84	3.82	3.24	3.15	3.15
700	6.74	6.71	6.72	4.79	4.93	4.67	3.81	3.92	3.83	3.32	3.12	3.09
750	6.24	6.04	6.03	4.52	4.56	4.58	3.76	3.72	3.71	3.17	2.95	2.98
800	5.74	5.71	5.69	4.31	4.32	4.32	3.61	3.59	3.56	2.91	2.90	2.90
850	5.22	5.19	5.17	4.12	4.02	3.98	3.36	3.34	3.32	2.97	2.73	2.71
900	4.87	4.86	4.82	3.85	3.83	3.70	3.17	3.16	3.16	2.59	2.58	2.58
950	4.44	4.40	4.37	3.64	3.59	3.59	3.04	3.01	2.98	2.46	2.44	2.44
1000	4.21	4.17	4.09	3.51	3.45	3.43	2.92	2.88	2.86	2.38	2.35	2.31
1050	3.90	3.85	3.78	3.27	3.23	3.19	2.76	2.73	2.69	2.22	2.21	2.20

The solution quality of the RMIP problem is generally satisfactory except for the instances of very restricted budgets. Though the DCMP method is a heuristic method and its solution's quality can not be theoretically guaranteed, its performance is empirically good enough with small optimality gaps. In particular, the solution quality of the DCMP method is empirically improved if the Scen2 is increased.

We have discussed the solution performance and quality of the DCMP. From the discussions, we know that the DCMP method is a promising direction for larger instances if the RMIP problem can be solved more efficiently.

Statistical interpretations to the optimal solutions

One challenge brought up by multistage stochastic programming (MSP) problems is how to interpret the scenario contingent solutions and how to implement the solutions in real life decision making. Unlike the decision tree examples taught in classrooms, which only consider a few simple "up or down" states, the solutions of the MEPPO problem involves complex decision variables and contingencies of higher dimensions along hundreds of thousands of sample paths. It is impractical to interpret each solutions stage by stage and path by path. We need a more concise and intuitive way to interpret the solutions.

In the remaining of this section, we take a statistical view to interpret the solutions based on the (conditional) expectation concept. We will only give examples for the three decisions, acquisition, appraisal, and drilling decisions.

Table 4-15 below lists the acquisition decisions for partial combinations of the three control factors. We can see that the oil field 2 is never selected in our project mix. We can find the reason from Table 4-4 and Table 4-5. This is mainly because the reservoir size is relatively too small and the COS is too low, compared to other oil fields under considerations. The Table 4-15 also tells us how project mix evolves as long as other control factors vary. Field 1 and 5 seem have higher priority over field 3 and 4 since whenever field 3 and/or field 4 is selected the field 1 and 5 must be selected as well.

Table 4-15 Acquisition decisions (partial)

Scen2	50												100											
ρ	0				0.2				0.4				0				0.2				0.4			
Budget (MM$) \ Fld	i1	i3	i4	i5	i1	i3	i4	i5	i1	i3	i4	i5	i1	i3	i4	i5	i1	i3	i4	i5	i1	i3	i4	i5
100	1			1	1			1	1			1	1			1	1	1		1	1	1		1
150	1			1	1			1	1	1		1	1	1		1	1	1		1	1	1		1
200	1	1		1	1	1		1	1	1		1	1	1		1	1	1		1	1	1		1
250	1	1		1	1	1		1	1	1		1	1	1		1	1	1		1	1	1		1
300	1	1		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		1	1	1	1	1
350	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
400	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
450	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
500	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
550	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
600	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
650	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
700	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
750	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
800	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
850	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
900	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
950	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1000	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1050	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

The following Table 4-16 presents the expected appraisal decisions cross all sample paths. Since the appraisal decision is binary, the table actually shows the probability to make appraisal decisions under the optimal investment policy. Since the appraisal decision is taken only if the field is proven “pay,” the values in the table should not exceed the corresponding COSs for selected oil fields. The COS values of all considering oil fields are listed at the bottom row of Table 4-4.

The following Table 4-17 lists the expected total number of wells to be drilled in selected fields, given the fields are “Pay”. One interesting finding to the table is the risk factor ρ has apparent impacts to the (empirical) drilling decisions only when the budget levels are low. The impacts diminish when the budget increase beyond certain level.

Table 4-16 Expected appraisal decisions (partial)

Scen2	50								100							
ρ	0				0.4				0				0.4			
Budget (MM$) \ Fld	i1	i3	i4	i5	i1	i3	i4	i5	i1	i3	i4	i5	i1	i3	i4	i5
100	0.557			0.182	0.547			0.191	0.6			0.183	0.546	0.188		0.168
150	0.59			0.227	0.568	0.254		0.195	0.595	0.321		0.218	0.6	0.286		0.218
200	0.614	0.361		0.233	0.614	0.399		0.219	0.614	0.382		0.223	0.614	0.392		0.232
250	0.6	0.399		0.241	0.614	0.399		0.233	0.614	0.417		0.24	0.614	0.417		0.232
300	0.614	0.42		0.244	0.607	0.421	0.209	0.233	0.614	0.401	0.26	0.232	0.614	0.408	0.26	0.232
350	0.614	0.432	0.209	0.241	0.614	0.432	0.209	0.241	0.614	0.432	0.266	0.241	0.614	0.432	0.26	0.241
400	0.614	0.441	0.229	0.251	0.614	0.441	0.236	0.251	0.614	0.441	0.268	0.251	0.614	0.441	0.26	0.251
450	0.614	0.441	0.258	0.251	0.614	0.441	0.243	0.251	0.614	0.441	0.282	0.251	0.614	0.441	0.266	0.251
500	0.614	0.441	0.268	0.251	0.614	0.441	0.281	0.251	0.614	0.441	0.282	0.251	0.614	0.441	0.282	0.251
550	0.614	0.441	0.268	0.251	0.614	0.441	0.281	0.251	0.614	0.441	0.282	0.251	0.614	0.441	0.282	0.251
600	0.614	0.441	0.268	0.251	0.614	0.441	0.281	0.251	0.614	0.441	0.289	0.251	0.614	0.441	0.284	0.251
650	0.614	0.441	0.279	0.251	0.614	0.441	0.281	0.251	0.614	0.441	0.289	0.251	0.614	0.441	0.297	0.251
700	0.614	0.441	0.282	0.251	0.614	0.441	0.281	0.251	0.614	0.441	0.305	0.251	0.614	0.441	0.305	0.251
750	0.614	0.441	0.289	0.251	0.614	0.441	0.289	0.251	0.614	0.441	0.314	0.251	0.614	0.441	0.314	0.251
800	0.614	0.441	0.289	0.251	0.614	0.441	0.289	0.251	0.614	0.441	0.314	0.251	0.614	0.441	0.314	0.251
850	0.614	0.441	0.289	0.251	0.614	0.441	0.289	0.251	0.614	0.441	0.314	0.251	0.614	0.441	0.314	0.251
900	0.614	0.441	0.289	0.251	0.614	0.441	0.289	0.251	0.614	0.441	0.314	0.251	0.614	0.441	0.314	0.251
950	0.614	0.441	0.289	0.251	0.614	0.441	0.289	0.251	0.614	0.441	0.314	0.251	0.614	0.441	0.314	0.251
1000	0.614	0.441	0.289	0.251	0.614	0.441	0.299	0.251	0.614	0.441	0.314	0.251	0.614	0.441	0.314	0.251
1050	0.614	0.441	0.289	0.251	0.614	0.441	0.289	0.251	0.614	0.441	0.314	0.251	0.614	0.441	0.314	0.251

Table 4-17 Expected aggregated drilling decisions conditional on “Pay” state (partial)

Scen2	50												100											
ρ	0				0.4				0.8				0				0.4				0.8			
Budget (\ Fld	i1	i3	i4	i5	i1	i3	i4	i5	i1	i3	i4	i5	i1	i3	i4	i5	i1	i3	i4	i5	i1	i3	i4	i5
100	47.6			14.9	46.0			14.9	30.7			7.1	39.8			14.0	29.1	7.4		13.5				
150	64.6			15.7	55.0	8.9		14.1	58.0	10.6		12.5	41.2	8.3		9.9	48.9	10.5		15.5	50.0	9.8		14.6
200	72.6	14.6		16.5	72.8	16.0		16.3	73.2	16.0		17.1	49.3	15.4		13.2	59.4	15.9		18.5	60.6	15.0		19.1
250	75.9	18.7		18.3	78.9	18.9		18.2	77.8	18.5		18.2	69.6	19.8		23.6	68.2	19.2		22.6	69.1	19.9		22.6
300	83.8	19.1		20.2	82.5	19.5	4.9	20.2	80.4	19.3	9.0	19.4	61.4	20.1	5.2	16.2	72.4	20.1	6.9	22.9	77.3	21.9		25.1
350	92.8	20.0	6.4	22.0	92.6	19.9	6.4	22.0	92.9	20.0	6.4	22.3	59.5	18.0	4.9	18.7	79.5	22.8	7.8	24.4	77.8	22.4	8.2	24.5
400	99.1	22.4	9.6	24.3	98.2	22.5	9.0	23.8	98.0	22.5	7.6	23.8	86.2	24.1	8.5	27.9	85.2	24.2	8.3	27.7	86.1	24.2	8.3	27.9
450	100.0	22.7	12.4	24.4	97.9	22.7	11.3	24.8	98.0	22.8	12.4	24.9	87.6	24.5	9.5	29.6	87.9	24.5	8.5	30.4	87.6	24.6	9.5	31.0
500	107.6	24.0	15.2	27.0	107.2	23.8	14.9	26.9	107.9	23.8	15.2	27.0	92.9	25.5	9.7	32.9	93.0	25.6	9.7	32.1	93.0	25.4	10.9	32.3
550	108.8	24.0	15.2	26.9	109.6	24.0	15.7	27.9	109.8	24.0	12.9	27.3	94.9	26.1	10.1	33.7	95.0	26.2	10.0	33.8	95.2	26.0	11.2	33.1
600	110.8	24.8	15.7	29.0	111.1	24.5	16.1	28.6	111.1	24.5	16.1	28.5	96.1	26.6	10.7	34.9	96.2	26.6	10.6	35.1	96.5	26.7	11.5	35.2
650	112.6	25.1	16.2	30.7	112.6	25.2	16.3	30.6	112.6	24.8	16.3	30.1	98.1	27.4	11.1	35.7	98.3	27.3	11.9	36.0	97.8	27.3	12.6	36.0
700	113.7	25.1	16.6	31.4	113.7	24.9	16.5	31.3	113.8	25.1	16.6	31.3	98.9	27.4	12.2	37.2	99.3	27.3	12.3	37.4	99.3	27.1	13.0	37.5
750	116.6	25.7	16.7	33.3	116.9	25.8	16.8	33.1	116.9	25.8	16.4	32.7	100.5	27.7	13.3	39.6	101.1	27.8	13.4	39.3	101.2	27.9	13.4	39.4
800	119.9	25.8	16.8	33.6	120.1	25.8	17.0	33.6	120.1	25.9	16.6	33.4	103.0	27.9	13.6	40.6	103.0	28.0	13.6	40.6	102.9	28.0	13.6	40.5
850	120.0	25.8	17.0	34.4	121.1	25.9	17.0	34.9	120.9	25.9	16.6	34.7	103.3	27.9	13.7	41.4	103.8	28.1	13.6	41.6	103.3	28.0	13.6	41.0
900	121.1	25.9	17.0	35.5	120.5	25.5	17.0	34.8	121.3	26.0	16.6	35.4	103.9	28.3	13.6	43.0	104.1	28.3	13.5	43.0	104.1	28.5	13.6	43.0
950	121.6	27.0	17.0	36.6	121.7	26.9	17.0	36.5	121.1	26.4	16.6	35.7	104.4	29.1	13.5	43.3	104.5	29.0	13.5	43.4	104.6	29.1	14.3	43.1
1000	121.4	26.7	17.0	36.8	121.3	26.8	17.1	36.7	122.2	26.9	16.7	36.7	104.5	28.9	13.7	44.6	104.8	29.2	13.7	44.3	104.4	29.1	14.5	44.3
1050	122.8	27.5	17.4	36.9	122.8	27.5	17.4	37.2	122.8	27.5	17.4	36.8	105.4	29.7	13.9	45.9	105.3	29.6	13.8	45.5	105.3	29.6	14.6	45.4

In order to have a better understanding about the drilling schedules, we can draw the histograms of the drilling decisions over production periods for each DCMP solution. The following Figure 4-7 gives an example what the drilling histograms would look like.

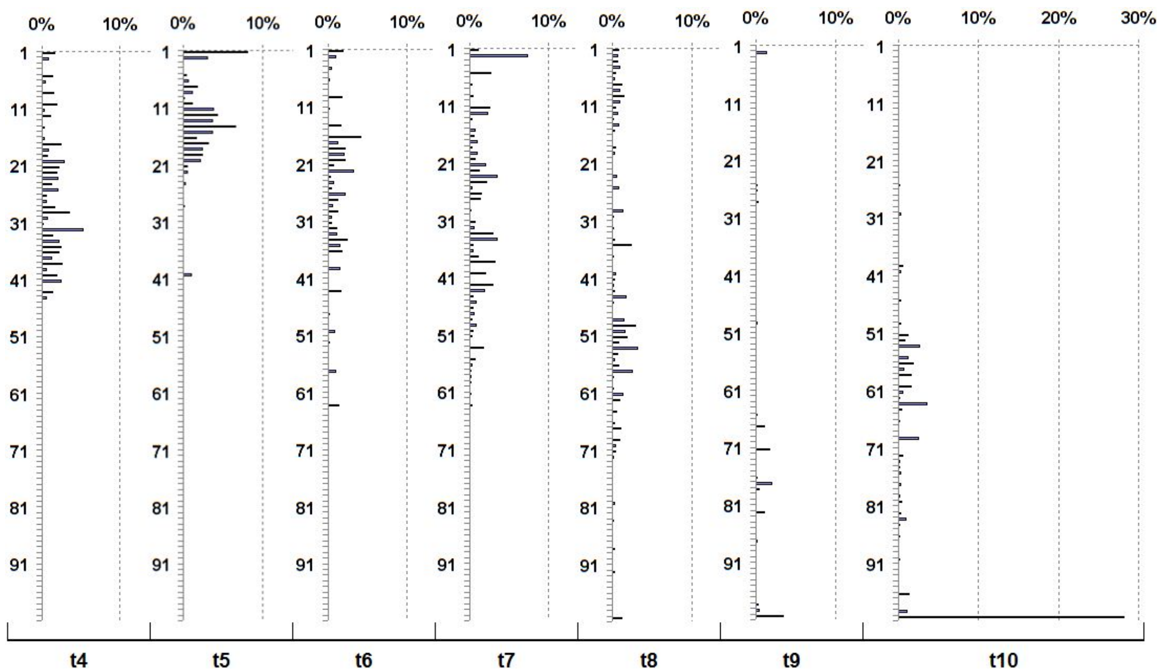


Figure 4-7 Histograms (conditioned on the "Pay" state) of the drilling decisions in oil field 1 over each production period, where *Scen2* = 50, ρ = 0.4, and *Budget* = 500 MM$. Vertical axes account for the number of wells to be drilled in a stage and horizontal axes for the frequency. (Note: zero drilling decisions have been excluded.)

For simplicity, we only show the histograms for oil field 1 with *Scen2* = 50, ρ = 0.4, and *Budget* = 500 MM$. We can draw histograms for other fields as well. The histograms summarize the drilling schedules in an intuitive way. The pattern shown in Figure 4-7 that less wells (measured by the median of each histogram) are drilled in earlier periods while more wells in later periods. This is reasonable due to the MEPPO properties and assumptions that we shall explain next. First, at the beginning, the budget is limited, while as production continues, the budget (the inventory cash position at each period) is accumulated from net revenues. Particularly, the MEPPO model allows revenues generated by one field to be invested on other fields. The dynamic budgeting and synergy effects together make it possible to drill more wells as production goes on. Second, we have assumed constant costs for both drilling and production and steadily growing oil prices. The deferral to drill a well has two complementary effects simultaneously, losing convenience yield and saving drilling and production costs. It is reasonable to think that there exits a break-even point with the former effect dominating the latter on one side and contrary on the other.

4.5 CONCLUSION

We have developed the multistage exploration and production (E&P) portfolio optimization (MEPPO) model. In the MEPPO model, multiple E&P projects are simultaneously managed to optimize an expected utility function which reflects the investor's preference toward the portfolio value and downside risk. The resulting MEPPO model is a large scale mixed integer program (MIP) with continuous variables to handle production forecasting data and various cash flows, binary variables to address capital expenditure decisions, and integer variables to schedule drilling plans. It is usually insufficient to simply call a commercial solver like CPLEX with default configurations to solve the MEPPO problem. Even to find a (non-trivial) integer feasible solution takes too much time. We developed three optimization based heuristic methods to expedite the solution procedure. Among them, a decomposition scheme is proposed to find good integer feasible solutions. Extensive numerical experiments have shown the effectiveness of the method. Empirically, an optimality gap of 5% is expectable for most instances and 20% for the most difficult instances with a very tight budget. To ensure the correctness of the model implementation, we propose a few verification methods to test the model from different aspects. We also provide general ideas how to interpret complex solutions of any multistage stochastic programming from a statistical view.

All above efforts have made it realistic to manage multiple E&P projects in a portfolio way. Since the MEPPO model is implemented in a flexible way and concrete data processes are separated from the model, more applications and enhancements are possible in future continuing research.

Chapter 5 Scenario Tree Generation

In this chapter, we develop a flexible framework to generate the forward looking scenarios for the MEPPO model presented in Chapter 4. The scenarios are organized in a tree structure to reflect how our present knowledge about uncertainty evolves over time. Over decades of research, there are abundant literatures on how to generate a scenario tree with required statistical properties for multistage stochastic programs, such as Dupacova et al. (2000), Hoyland and Wallace (2001), and Heitsch and Romisch (2007). Besides the conventional requirements on statistical quality to generate scenario trees, we pay special attentions to the following two issues while designing our scenario generator.

First, E&P projects are long-term investments and we have diverse opportunities to learn and make educated decisions. We identify two types of statistical learning and incorporate them in the scenario generation to obtain better estimation for the performance of E&P projects. The two types of learning are based on two respective dependences. The first learning is called binary learning and captures the inter-project dependence. For example, the success to find commercial oil in one reservoir may improve the likelihood to find commercial oil in nearby reservoirs. Such dependences are modeled by finding a full joint distribution about all reservoirs' exploratory outcomes, whether there is commercial oil or not. The second learning is sequential linear least-squares learning and addresses the intra-project dependence, the dependence among (geological) parameters of the same reservoir. It progressively improves estimations about the reservoir's productivity using a sequence of observed data. In a special case where individual multi-normal random vectors jointly form "larger" multi-normal random vectors, the sequential least-squares learning degenerates to the sequential Bayesian learning since the best least-squares estimator of any multi-normal distribution has a linear form and coincides with the linear least-squares estimator. We provide theoretical supports to the traditional view that the sequential learning does progressively shrink the spreads of the updated distributions.

Both types of statistical learning rely on statistical dependence and observations. Given two dependent random vectors, the observed value of one vector provides statistical information to better estimate the value of the other. There are different but consistent ways to measure dependence, such as correlation (or covariance), marginal and conditional probabilities, and (partial or full) joint distributions. Among those methods, correlations are the least powerful and only captures linear statistical dependence. Full joint distributions are the most powerful and capture complete statistical dependences. Marginal and conditional probabilities together are in the between of the other two. However, information can not be acquired for free. The difficulty and (computational and monetary) costs to obtain them are just inversely proportional to their power, i.e., full joint distributions are the most expensive and difficult to obtain while correlations are just the contrary.

Second, we want to have certain level of control to trade the accuracy of the scenario tree off its size. One popular way to solve a multi-stage stochastic program (or a more general stochastic dynamic program) is to discretize both the time and the state spaces, then approximate the underlying stochastic process with a multi-stage scenario tree, and finally reformulate the problem as an equivalent deterministic problem. However, the discretization and the multistage tree structure usually result in very large problems. The computational challenges brought by the exponential growth of the scenario tree dominate the presence of discrete decision variables, see Shapiro (2006). To get solvable models, we have to limit the growth of the scenario tree by sacrificing statistical accuracy to some extent. A popular method to achieve this goal developed over the past ten years is to perform scenario reduction such that the reduced scenarios provide good approximation to the original ones with respect to some probability metrics. The basic idea to perform scenario reduction is a two-step procedure. First, the original distribution is replaced by a discrete distribution over a large number of scenarios. Then, a small subset of scenarios is selected and reassigned with a normalized distribution such that the reduced sample is close to the sample before reduction according to the probability metric. We applied two different scenario reduction methods to the two parts of the scenario tree generated by two different types of distributions.

This chapter is organized as follows. In §5.1, we introduce the structure of the scenario tree and two types of dependences and associated respective learning. In §5.2, we introduce a way to model the first type of learning using entropy-maximization and present an idea to reduce scenario based on empirical experiments. In §5.3, we introduce the second type of learning by sequential observations and updating. Both types of scenario reductions work together to make the size of the full scenario tree controllable. In §5.4, we provide numerical experiments to test the scenario generation and reduction codes. We also discuss advantages and weakness of our methods, propose other methods, and discuss extensions.

5.1 Structure of The Scenario Tree

Compared to most scenario trees reported in literature, the scenario tree for our E&P portfolio optimization model is more complex since it carries information from different sources and from multiple oil fields. Each branch out of a node in the scenario tree represents one realization of a vector of random parameters, which describe the geological properties of all fields. For simplicity, we break the complete scenario tree into two independent parts, called the first subtree and the second subtree, according to their disjoint timing and disparate nature of uncertainty. We will generate them separately based on different theories and then combine them together by attaching a replica of the second subtree to each leaf node of the first subtree to recover the complete scenario tree.

The first subtree describes the probabilistic information of the combinatorial "Pay/No pay" state of all fields. It is a single-phase tree with $2^{|\mathbb{I}|}$ leaf nodes, where $\mathbb{I}$ is the set of fields. The first subtree involves the uncertainty around the combinatorial binary information about the existence of commercial oil in all reservoirs. For each individual reservoir, this information represents the two possible states, either Pay or No Pay. Pay means that there exits commercial oil in a reservoir and no pay means the contrary. This information is resolved only if the field has been acquired and some preliminary exploration has been carried out to identify the commercial oil. Before acquisition and pre-exploration, we only have limited information about the probability of pay, known as the chance of success (COS). Since we have to make the acquisition decision with this limited information, it is crucial to get accurate estimation about the COS. In next section, we will apply a method called optimal binary learning to make educated estimation about the combinatorial information of COS of all fields. Figure 5.1 illustrates the first subtree for the "Pay/No Pay" information of three projects. The detailed explanations about the notation of this example are in §5.2.

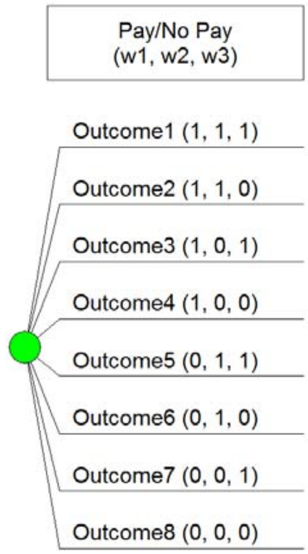


Figure 5-1: The first subtree: the Pay/No pay information of a portfolio with 3 projects. E.g. the third outcome (w1, w2, w3) = (1, 0, 1) corresponds the state that only the two reservoirs other than reservoir 2 are confirmed to contain commercial oil.

The second subtree captures how we sequentially refine our belief about the reservoirs' performance as investment activities progress. We model the belief updating with a sequential least-squares learning process. In Chapter 4, we use three-point approximation as an example to approximate the distribution of random parameters of single project since it is easier to understand. However, the three-point approximation applied in multi-project, multivariate, and multistage cases will unavoidably result in a fast growing scenario tree. Therefore, instead of using the three-point approximation, we propose an adaptive Monte Carlo simulation procedure to recursively generate the second scenario subtree, as shown in Figure 5.2.

The content of the second subtree is contingent on the outcome of the first subtree. Intuitively, the second subtree contains information about a reservoir only if the reservoir contains commercial oil, or its state is Pay, and the reservoir is acquired to explore. Otherwise, no more information about the reservoir will be obtained. To enforce the contingence and makes the resulting scenario tree consistently model our knowledge, we can disable the data on the second subtree for projects whose states on the first subtree status is zero (No Pay).

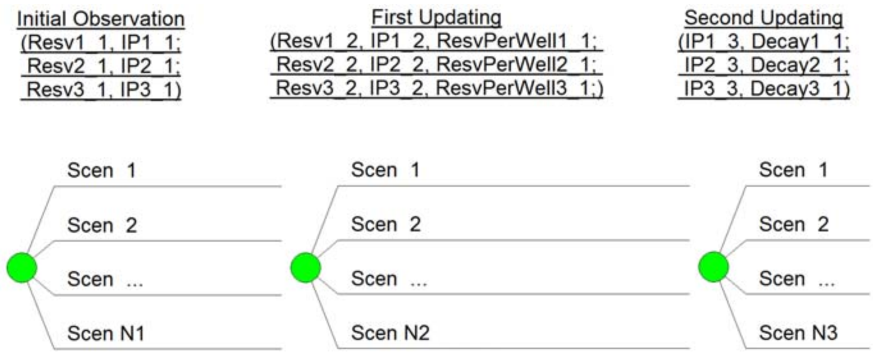


Figure 5-2: The second part of the scenario tree, given all reservoirs containing commercial oil: the 3-phase updating of a portfolio with 3 projects, where N1, N2, and N3 are the number of multivariate samples to be drawn at each learning stage.

Therefore, we can combine the separately generated information to represent the full knowledge of we. The separation makes the scenario generation easy to implement since we can handle different types of random variables separately. In section 5.2 and 5.3, we introduce methods to generate the two parts of scenario tree.

However, the separation also brings a challenge to us. The combination of the two separate trees usually results in large scale problems. For example, the theoretical number of the terminal nodes of the combined scenario tree based on Figure 5.1 and Figure 5.2 is $2^{|\mathbb{I}|} \cdot N_1 \cdot N_2 \cdot N_3$, where $|\mathbb{I}|$ is the number of projects.

5.2 LEARNING FROM INTER-PROJECT DEPENDENCE

In this section, we model the inter-project dependence by finding the full joint probabilities of the binary state "Pay/No Pay" across projects. We apply the optimal binary learning method of Bickel and Smith (2006) to recover this full joint distribution based on marginal and pairwise conditional probabilities, which are relatively easier to obtain. The derived joint probabilities will be attached to the corresponding branches of the first subtree, as shown in Figure 5.1. However, the catch is that the number of branches grows exponentially in the number of projects. We propose a simple approach to prune the branches whose probabilities are believed insignificant in magnitude. From our numerical experiments, the reduced first subtree grows slowly enough and allows us to solve problems of practical size.

5.2.1 Optimal binary learning

The optimal binary learning is a method to recover full joint probability for a random binary vector, based on its marginal and pairwise joint probability assessments. The idea is to choose a probability measure defined on the binary vector space to maximize an entropy measure such that the chosen probability is consistent with the marginal and pairwise assessments. To make the dissertation self-contained, we provide a brief review to the method. More information about the entropy method please refer to Luenberger (1984), Miller and Liu (2002), and Bickel and Smith (2006).

Consider $n = |\mathbb{I}|$ E&P projects in the portfolio. Let binary vector $\mathbf{w} \in \mathbb{B}^n$ to denote the joint "Pay/No pay" status of the *n* projects, where $\mathbb{B} = \{0,1\}$ and $|\mathbb{B}^n| = 2^n$. Each component w_i of the vector is a binary random variable to indicate where reservoir *i* contains oil (if $w_i = 1$) or not (if $w_i = 0$). The probability $p_i = \mathbb{P}(w_i = 1)$ is called the chance of success (COS) of reservoir *i*. Therefore, each realization of the random vector **w** tells us one combination of the "Pay/No pay" state of all projects. For reservoirs from the same geological area, the "Pay/No pay" outcomes of some of them may enhance our belief about the others. We are interested in modeling such inter-project dependence due to geological correlation.

We describe this statistical inter-project dependence by a joint probability $\pi^*(\mathbf{w}) \in \Pi$, where $\Pi := \left\{\pi : \mathbb{B}^n \to [0,1], \sum_{\mathbf{w}\in\mathbb{B}^n} \pi(\mathbf{w}) = 1\right\}$. The optimal binary learning method is to find a joint probability to solve the following constrained entropy maximization problem,

$$\min_{\pi\in\Pi} \quad \sum_{\mathbf{w}\in\mathbb{B}^n} \pi(\mathbf{w}) \ln\left(\frac{\pi(\mathbf{w})}{\pi_0(\mathbf{w})}\right) \tag{5.2.1}$$

$$\text{s.t.} \quad \mathbb{E}_\pi\left[\Omega_0(\mathbf{w})\right] = 1 \tag{5.2.2}$$

$$\mathbb{E}_\pi\left[\Omega_i(\mathbf{w})\right] = p_i, \ \forall i \tag{5.2.3}$$

$$\mathbb{E}_\pi\left[\Omega_{ij}(\mathbf{w})\right] = p_{ij}, \ \forall i, j \text{ and } j \le i, \tag{5.2.4}$$

where the pairwise joint probability p_{ij} and conditional probability $p_{j|i}$ are defined as $\mathbb{P}(w_i = 1, w_j = 1)$ and $\mathbb{P}(w_j = 1 \mid w_i = 1)$, respectively. Furthermore, $p_{ij} = p_{j|i} \cdot p_i$. The left hand side of equations (5.2.2) ~ (5.2.4) rely on the following three operators: (1) $\Omega_0(\mathbf{w}) = 1$, $\forall \mathbf{w} \in B^n$; (2) $\Omega_i(\mathbf{w}) = w_i$, $\forall \mathbf{w} \in B^n$, $\forall i$, and (3) $\Omega_{ij}(\mathbf{w}) = w_i \cdot w_j$, $\forall \mathbf{w} \in B^n$, $\forall i, j$. The equations of (5.2.2) ~ (5.2.3) are essentially the moment matching constraints and simply ensures that the chosen probability measure is consistent to given information up to the cross moments.

This problem is a convex program since the objective function is convex and all constraints are linear. Therefore, a local solution is also a global solution. Besides, any optimal solution is strictly positive due to the unlimited penalties imposed by the log functions in (5.2.1) if otherwise. So we can drop the nonnegativity constraints required by any valid distribution measure.

The objective function (5.2.1) defines the Kullback-Liebler (KL) distance relative to a reference distribution π_0,

$$D(\pi, \pi_0) = \sum_{\mathbf{w} \in B^n} \pi(\mathbf{w}) \ln\left(\frac{\pi(\mathbf{w})}{\pi_0(\mathbf{w})}\right). \tag{5.2.5}$$

It measures the informational distance between a target distribution π and the reference distribution. In information theory, a distribution π^* minimizes (5.2.5) is said to be "least informative," or equivalently "maximally uncertain" relative to π_0. The naturally occurring probability is always characterized by the least informative distribution. Therefore, this approach generates a conservative distribution using given information.

Particularly, when π_0 assumes mutual independence as below,

$$\pi_0(\mathbf{w}) = \prod_i (p_i)^{w_i} (1 - p_i)^{1-w_i}, \tag{5.2.6}$$

the KL distance actually measures the strength of dependence in the joint distribution π. So the model aims to find a distribution of the minimal amount of dependence within all consistent distribution measures.

5.2.2 Solution and a crude reduction

The optimal binary learning model (5.2.1) ~ (5.2.4) is a convex program and its solution has nice analytical structure. Bickel and Smith (2006) have proposed two different ways to solve it. However, the solution is a full joint distribution over all 2^n distinct elements of the Cartesian product $\mathbb{B}^n$. The solution(s) of the binary learning has a property that all components are strictly positive. So the resulting scenario subtree corresponding to the COS states will grow very fast. Fortunately, from our numerical experiments, we found that a few elements have probabilities which are significantly larger than the remaining majority of elements. So we can scan the solution to find some good cutoff point to truncate the full joint distribution. In our scenario generation code, we sort and scan the probabilities from high to low. We choose the cutoff point to be the one whose probability is 1,000 times of its successor which has the highest probability among the remaining sample space. For our five project example, 13 elements out of the full population of size $2^5 = 32$ have significantly higher probabilities than others. Our experiments show that the reduced probability space has good approximation to the marginal probabilities and most pair-wise joint probabilities. However, some pair-wise joint probabilities do show poor approximations. We shall leave the search for better and more comprehensive reduction algorithms as future research topics.

From our research, we realize that dimensional reduction and scenario reduction are playing increasingly more important roles in multistage stochastic optimization. They deserve more attentions in future research.

5.3 LEARNING FROM INTRA-PROJECT DEPENDENCE

Once a reservoir is verified to contain commercial oil, we will make a sequence of decisions to construct production facility and make drilling plans. These decisions are contingent on the reservoir performance, oil price, operational costs, and so on. All those factors can be sources of uncertainty to impact the decisions and the project value. Of particular interest to us is the reservoir performance, measured by geological properties like reserve size and production rates. We will ignore the market risk and stick to a fixed oil price or a fixed forward curve. We also assume all cost terms are deterministic. Such simplifications are reasonable for E&P projects since in the early stages of a field, the geological risk dominates the fluctuations in oil price and costs. Moreover, an oil field starts to generate revenues only if it proceeds to the production phase in from a few years up to ten years after the first wildcat drilling.

The second subtree models how we progressively refine our belief on reservoir parameters, based on seismic data, core samples, and well logs. Those parameters will be used to estimate the reservoir commerciality, such as original oil in place (OOIP), productivity, and recovery rate. We will use the tank model to forecast the field productivity. However, the tank model introduced in §2.2 directly deals with geological parameters and makes the production rates highly nonlinear in well numbers, which are main decisions variables in our E&P portfolio model. A large number of binary variables will be incurred to linearize the tank model. To avoid the computational burden, we assume the production rates of all wells in the same field can be forecasted by the same exponential decline curve.

5.3.1 Simplified tank model

The tank model introduced in §2.2 ignores spatial variations and assumes geological parameters to be unanimous everywhere in the reservoir. In our E&P project portfolio model, we further assume wells are drilled in proper spacing such that each well can be roughly treated as being isolated from each other. So the depletion of one well will not affect the productivity of other wells. Wells drilled at any time in a reservoir have the same production decline curve. Given an estimated recoverable reserve, we may drill enough wells over time so that the accumulated productions achieve the target. The number of wells to be drilled can be roughly estimated by the division of the (estimated) total reserves (R) and the (estimated) reserves per well (WR).

The tank model is determined by four parameters, well reserve (WR), initial production rate (Q_0), decay rate (λ), and the economic limit (Q_{eco}). Those parameters must satisfy a consistence condition that $WR = \left(Q_0 - Q_{eco}\right)/\lambda$, which is bounded from above by Q_0/λ. So as long as $WR \le Q_0/\lambda$ is guaranteed, if any three out of the four parameters are known, the fourth parameter is uniquely determined. So the reservoir development and production strategies will depend on four random parameters: reservoir reserve (R), reserve per well (WR), initial well production rate (Q_0), and decay rate(λ). They summarize sufficient geological information contained in more fundamental raw data, such as porosity and permeability, to forecast the reservoir performance. We can treat the four parameters as a vector-valued sufficient statistic for our tank model.

5.3.2 Sequential statistical learning

We propose a statistical learning process to describe how the joint distribution of the four parameters (reservoir reserve, reserve per well, initial well production rate, and decay rate) is sequentially updated during the development and production. The updated distributions will be used to recursively generate the second subtree using Monte Carlo simulation techniques. In such a way, we can generate a scenario tree which takes into account statistical learning effects. We will show that the resulting scenario tree does represent a stochastic process with a shrinking variance in each component over time. If the learning keeps on going infinitely, the process eventually converges to a limiting point estimate of the vector of random parameters almost surely.

At different stages, some of the parameters are observed and we can use observations as indirect evidence to infer unresolved information and refine our belief from time to time. For example, appraisal wells and early production data provide probabilistic information about the reserve amount and initial production rates. As production continues, we have more data sufficient to estimate the decay factor and improve earlier beliefs. The learning process is guided by the linear least-squares estimation for parameters of general distributions and by the Bayesian updating principle in the Gaussian case. We shall see that our sequential learning procedure coincides with the Kalman filtering. From this perspective, our E&P portfolio model is essentially a multistage stochastic programming approach to solve the optimal control and estimation problem for the risk management of E&P project portfolios.

Once we have the stochastic process, we can recursively generate the scenario tree with a breadth-first search procedure as follows. At any non-terminal node of the tree, we generate branches fanning out from the current node in a three-step procedure. First, we update the distribution based on the nodal resolved information. Second, we draw a large set of equal-likely samples from the updated distribution of the current node. Third, we apply a scenario reduction procedure to reduce the large sample set to a small number of branches out of the node, with each branch associated with an adjusted conditional probability, also called one-step transition probability. We concentrate on the first two steps in the remaining of this subsection. In next subsection, we will introduce the idea of scenario reduction and apply a popular scenario reduction code, GAMS/Scenred, to control the growth of the scenario tree.

Fundamental statistical updating rules

The sequential updating process relies on the following two fundamental statistical learning rules, Proposition 5.1 and Proposition 5.2, based on Bertsekas (2007). The idea is the same as the linear least-squares regression, simply to estimate a random vector using observations of another random vector, given they are correlated.

Proposition 5.1 deals with multinormal distributions and coincides with the Bayesian updating principle. Proposition 5.2 handles more cases of arbitrary distributions.

We setup some frequently used notations first. Given random vectors $X \in R^n$ and $Y \in R^m$, we denote $\mu_X = \mathop{\mathbb{E}}_X[X]$ and $\mu_Y = \mathop{\mathbb{E}}_Y[Y]$ as their respective means, $\Sigma_{XX} = \mathop{\mathbb{E}}_X\left[(X-\mu_X)(X-\mu_X)'\right]$ and $\Sigma_{YY} = \mathop{\mathbb{E}}_Y\left[(Y-\mu_Y)(Y-\mu_Y)'\right]$ their variance-covariance matrixes, and $\Sigma_{XY} = \Sigma_{YX}{}' = \mathop{\mathbb{E}}_{X,Y}\left[(X-\mu_X)(Y-\mu_Y)'\right]$ the covariance matrix.

Proposition 5.1 (**Multinormal Bayesian Updating**) Given a multinormal random vector $\theta \sim N(\mu, \Sigma)$, we can partition it into two sub-vectors, *X* and *Y*, such that $\theta = (X', Y')'$ and

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{pmatrix}\right). \tag{5.3.1}$$

Without observations of *Y*, the estimate of *X* is μ_X with error covariance matrix Σ_{XX}. Given an observation *y*, the least-squares estimate of *X* is the conditional expectation,

$$\mu_{X|y} = \mathbb{E}\left[X \middle| Y = y\right] = \mu_X + \Sigma_{XY}\Sigma_{YY}^{-1}(y - \mu_Y) \tag{5.3.2}$$

with shrunk error covariance matrix

$$\hat{\Sigma}_{XX} = \mathop{\mathbb{E}}_{X,Y}\left[\left(X - \mu_{X|Y}\right)\left(X - \mu_{X|Y}\right)'\right] = \Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}. \qquad \square \tag{5.3.3}$$

Proposition 5.2 (**Linear Least-Squares Estimation**) For two random vectors *X* and *Y* of arbitrary distributions, given their covariance matrix Σ_{XY}, we can obtained the linear least-squares estimate and error covariance matrix of *X* based on observation *y* as

$$\widehat{X}(y) = \mu_X + \Sigma_{XY}\Sigma_{YY}^{-1}(y - \mu_Y), \tag{5.3.4}$$

$$\widehat{\Sigma}_{XX} = \underset{X,Y}{\mathbb{E}}\left[\left(X - \widehat{X}(Y)\right)\left(X - \widehat{X}(Y)\right)'\right] = \Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}. \tag{5.3.5}$$

□

Proposition 5.3 (**Orthogonal Projection Principle**) Given the linear least-squares estimate $\widehat{X}(y)$, the estimation error $X - \widehat{X}(Y)$ is uncorrelated with both Y and $\widehat{X}(Y)$,

$$\underset{X,Y}{\mathbb{E}}\left[Y\left(X - \widehat{X}(Y)\right)'\right] = 0 \quad \text{and} \quad \underset{X,Y}{\mathbb{E}}\left[\widehat{X}(Y)\left(X - \widehat{X}(Y)\right)'\right] = 0.$$

□

Proposition 5.4 Given two multinormal random vectors $X \sim N(\mu_X, \Sigma_{XX})$ and $Y \sim N(\mu_Y, \Sigma_{YY})$, the combined vector $\theta = (X', Y')'$ is not necessary to be multi-normally distributed. However, if they are correlated with a known covariance matrix Σ_{XY}, we still can improve the estimate of the population mean μ_X by its linear least-squares estimate as (5.3.4) based on observation *y* of *Y*. The covariance matrix Σ_{XX} can be estimated by the estimate error covariance matrix (5.3.5). □

More detailed discussions please refer to Bertsekas (2007). The next propositions take care of statistical learning from multiple observations. The observations may or may not be correlated. Although the timing of the observations is immaterial, we tend to treat the observations coming in sequence. Proposition 5.5 sets the foundation to sequentially update an existing linear least-squares estimate by incrementally incorporate new observations. The resulting estimate at the final step is the same as if the estimate is obtained by regressing over all aggregated observations. We will extend it in Proposition 5.6 to show a recursive form which facilitates our sequential updating. The propositions serve as the foundation of the Kalman filtering and allow us to simplify the learning calculations.

Proposition 5.5 Given a new random vector *Z*, let it serve as a new measurement to Proposition 5.2. Given two covariance matrixes Σ_{XZ} and Σ_{YZ}, we can form a better linear least-squares estimate $\widehat{X}(y,z)$ by incorporating both values y and z in the following two equivalent ways.

(a) The first way is to combine the two measurements into an aggregated random vector $U=(Y',Z')'$ with observation $u=(y',z')'$. The mean of U is $\mu_U=\left(\mu_Y{}',\mu_Z{}'\right)'$ and the variance-covariance matrix is $\Sigma_{UU}=\begin{pmatrix}\Sigma_{YY} & \Sigma_{YZ}\\ \Sigma_{ZY} & \Sigma_{ZZ}\end{pmatrix}$. The covariance matrix between X and U is $\Sigma_{XU}=\begin{pmatrix}\Sigma_{XY} & \Sigma_{XZ}\end{pmatrix}$. By Proposition 5.2, we have the linear least-squares estimate and error variance matrix as

$$\widehat{X}(y,z)=\widehat{X}(u)=\mu_X+\Sigma_{XU}\Sigma_{UU}{}^{-1}(u-\mu_U), \tag{5.3.6}$$

$$\widehat{\Sigma}_{XX}=\mathop{\mathbb{E}}_{X,U}\left[\left(X-\widehat{X}(U)\right)\left(X-\widehat{X}(U)\right)'\right]=\Sigma_{XX}-\Sigma_{XU}\Sigma_{UU}{}^{-1}\Sigma_{UX}. \tag{5.3.7}$$

In the case when *Y* and *Z* are uncorrelated and therefore $\Sigma_{ZY}=\Sigma_{ZY}{}'=0$, we can simplify the (5.3.6) and (5.3.7) as

$$\widehat{X}(y,z)=\widehat{X}(y)+\widehat{X}(z)-\mu_X, \tag{5.3.8}$$

$$\widehat{\Sigma}_{XX} = \Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}{}^{-1}\Sigma_{YX} - \Sigma_{XZ}\Sigma_{ZZ}{}^{-1}\Sigma_{ZX}\,. \tag{5.3.9}$$

(b) The second way is to update the estimate $\widehat{X}(y)$ by incorporating new information contained to form the new estimate $\widehat{X}(y,z)$. Observing value *y*, we can compute estimate $\widehat{X}(y)$ and $\widehat{Z}(y)$. Since $\widehat{Z}(y)$ is linear in *y*, $\widehat{X}(y,z) = \widehat{X}\left(y, z-\widehat{Z}(y)\right)$ by the definition and optimality of the linear least-squares optimization problems. Moreover, we can show *y* is uncorrelated with $z-\widehat{Z}(y)$ by direct computation. Therefore, applying (5.3.8) and (5.3.9), we have the following recursive formulae,

$$\widehat{X}(y,z) = \widehat{X}(y) + \widehat{X}\left(z-\widehat{Z}(y)\right) - \mu_X\,, \tag{5.3.10}$$

$$\widehat{\Sigma}_{XX} = \Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}{}^{-1}\Sigma_{YX} - \widehat{\Sigma}_{XZ}\widehat{\Sigma}_{ZZ}{}^{-1}\widehat{\Sigma}_{ZX}\,. \tag{5.3.11}$$

Where

$$\widehat{\Sigma}_{XZ} = \mathop{\mathbb{E}}_{X,Y,Z}\left[\left(X-\mu_X\right)\left(Z-\widehat{Z}(Y)\right)'\right] = \Sigma_{XZ} - \Sigma_{XY}\Sigma_{YY}{}^{-1}\Sigma_{YZ}\,,$$

$$\widehat{\Sigma}_{ZZ} = \mathop{\mathbb{E}}_{Y,Z}\left[\left(Z-\widehat{Z}(Y)\right)\left(Z-\widehat{Z}(Y)\right)'\right] = \Sigma_{ZZ} - \Sigma_{ZY}\Sigma_{YY}{}^{-1}\Sigma_{YZ}\,. \qquad \square$$

Parameter and observation processes

In petroleum industry, it is common to assume the four parameters (reservoir size, reserve per well, initial well production rate, and decay rate) to follow a multi-lognormal distribution, i.e., the logged parameters follow a multinormal distribution. There exist closed-form transformations between the original data's variance-covariance matrix and the logged data's variance-covariance matrix. The transformations guarantee that if we take exponentials to the samples drawn from the multi-normal distribution, the resulting data's variance-covariance matrix is consistent to that of the original data. Moreover, the log function is a one-to-one mapping which is strictly increasing. Guaranteed by the equivalence of the transformations, we will concentrate on the logged parameters.

Let column vector $\theta \in R^4$ denote the logged data of the four parameters, i.e., $\theta = (\theta_1, \theta_2, \theta_3, \theta_4)' = (\log(R), \log(WR), \log(Q_0), \log(\lambda))'$. The vector θ follows a multinormal distribution $N(\mu, \Sigma)$. When the distribution of θ evolves over discrete time points, we use the vector process $\{\theta_t, t = 0,1,\cdots\}$ to represent the generated discrete time stochastic process, where $\theta_t = (\theta_{1t}, \theta_{2t}, \theta_{3t}, \theta_{4t})'$. There exists a multinormal distribution $N(\mu_t, \Sigma_t)$ for each time point *t* with $\mu_t \in R^4$ and $\Sigma_t \in R^{4\times 4}$, where Σ_t is assumed to be a symmetric positive definite matrix, such that $\theta_t \sim N(\mu_t, \Sigma_t)$. The true values of the parameters μ_t and Σ_t are never known. We just try as much as we can to improve the estimate of them. At the beginning, we only have an inaccurate prior estimation about those parameters, based on experts' opinions and preliminary surveys. We will try to refine the evolution of θ_t to find best possible estimation about the parameters by incrementally incorporating the latest resolved information.

Associated with the process $\{\theta_t, t = 0,1,\cdots\}$ is a sequence of observations of one or more components of the vector θ during the development of the reservoir. To describe the observations, we partition the vector θ_t into two disjoint sub-vectors X_t and Y_t such that Y_t represents the components whose values are observable at time *t* and X_t represents the remaining components. Therefore, the sequence $\{Y_t, t = 0,1,\cdots\}$ represents the observable components. Depending on which components of θ_t are observable at time *t*, the dimensions of both vectors X_t and Y_t may change from time to time, but the total of their dimensions equals to the dimension of θ_t, which is 4.

Since the dynamics of the four parameters is driven by the same geological conditions, it is likely that there exits serial correlation in the process $\{\theta_t, t = 0,1,\cdots\}$.

The serial correlation is the correlation of the random vector θ with itself over successive time intervals. It conveys useful information to predict the future performance of the four parameters, based on their current values. We assume the serial covariance matrix $\Sigma_{\theta_s Y_t} = \underset{Y_t,\theta_s}{\mathbb{E}}\left[\left(\theta_s - \mu_{\theta_s}\right)\left(Y_t - \mu_{Y_t}\right)'\right]$ ($s > t$) are known as part of the prior belief. So we can use the observation y_t to update the belief about θ_s, where $s > t$.

When Y_t and θ_s together follow a multinormal distribution, we can apply the following Proposition 5.6 (the Bayesian updating principle) to compute the updated distribution of θ_s conditional on realizations of Y_t in closed-form. Otherwise, the Bayesian rule doesn't apply and it is difficult to get the closed-form conditional distribution of θ_s. In this case, we will rely on sampling and the linear least-squares estimate $\widehat{\theta}_s\left(y_t\right)$ to improve the point estimate $\overline{\theta}_s$ of μ_s. In either way, the updating formulae of the parameters of the distribution of θ_s are the same, although the background statistical meanings are different. The Bayesian update achieves the best least-squares estimate, while the linear least-squares method only finds the best estimate among the family of linear estimates.

Sequential updating and scenario generation

Our scenario generator is a recursive procedure which involves three activities, learning, sampling, and reduction. The efficient implementation of statistical learning relies on the following Proposition 5.6. It is an application of the Proposition 5.5 in the multistage situation. With the proposition, we can represent the updating formulae in a recursive form, which fits into our scenario generation code very well.

Consider a stream of observations $\{y_1, \cdots, y_t\}$ up to time *t*, we introduce a vector $I_t = \left(y_1, \cdots, y_t\right)$ to read them column-wisely. The column vector I_t denotes the observation history up to and including time *t*. In recursive form, $I_0 = \varnothing$ and $I_t = I_{t-1} \cup \{y_t\}$. We denote $\widehat{\theta}_{s|t-1} = \widehat{\theta}_s(I_{t-1})$ and $\widehat{\Sigma}_{s|t-1} = \underset{\theta_s, \{Y_u\}_{u=1}^{t-1}}{\mathbb{E}}\left[\left(\theta_s - \widehat{\theta}_{s|t-1}\right)\left(\theta_s - \widehat{\theta}_{s|t-1}\right)'\right]$ as the updated prior estimate of $\left(\overline{\theta}_s, \Sigma_s\right)$ based on the observation I_{t-1}, for $s > t$. We start the learning procedure with a prior consisting of $\left\{\left(\widehat{\theta}_{t|0}, \widehat{\Sigma}_{t|0}\right) = \left(\overline{\theta}_t, \Sigma_t\right), t = 1, \cdots, T-1\right\}$ and $\left\{\Sigma_{\theta_s Y_t}, \forall 1 \le t < s \le T\right\}$. We want to incrementally improve the estimation on $\left(\overline{\theta}_s, \Sigma_s\right)$ using all correlated observations prior to *s*. Before time *t*, we have observations $I_{t-1} = \left(y_1, \cdots, y_{t-1}\right)$ and have used them to get the updated parameters $\widehat{\theta}_{s|t-1}$ and $\widehat{\Sigma}_{s|t-1}$, for *s* > *t*-1. When a new observation y_t comes at time *t*, we can keep on updating the parameters to $\widehat{\theta}_{s|t}$ and $\widehat{\Sigma}_{s|t}$. The Proposition 5.6 tells how to carry out the updating.

Proposition 5.6 Given observations $I_{t-1} = \left(y_1, \cdots, y_{t-1}\right)$, we can apply the following recursive procedure to calculate the linear least-squares estimate $\widehat{\theta}_{s|t-1} = \widehat{\theta}_s\left(I_{t-1}\right)$ and the error variance matrix $\widehat{\Sigma}_{s|t-1} = \underset{\theta_s, \{Y_u\}_{u=1}^{t-1}}{\mathbb{E}}\left[\left(\theta_s - \widehat{\theta}_{s|t-1}\right)\left(\theta_s - \widehat{\theta}_{s|t-1}\right)'\right]$ of $\left(\overline{\theta}_s, \Sigma_s\right)$, for *s* = *t-1*, *t*, …, $T-1$. If a new observation y_t comes at time *t*, we can update the estimates from ($\widehat{\theta}_{s|t-1}$, $\widehat{\Sigma}_{s|t-1}$) to ($\widehat{\theta}_{s|t}$, $\widehat{\Sigma}_{s|t}$), for $s = t+1, \cdots, T-1$, according to the following formulae

$$\widehat{\theta}_{s|t} = \widehat{\theta}_{s|t-1} + \Sigma_{\theta_s Y_t} \widehat{\Sigma}_{Y_t Y_t|t-1}^{-1}\left(y_t - \widehat{Y}_{t|t-1}\right), \tag{5.3.12}$$

$$\widehat{\Sigma}_{s|t} = \widehat{\Sigma}_{s|t-1} - \Sigma_{\theta_s Y_t} \widehat{\Sigma}_{Y_t Y_t|t-1}^{-1} \Sigma_{Y_t \theta_s}. \tag{5.3.13}$$ □

[Proof] At *t-1*, we partition the vector θ_t into $\begin{pmatrix} X_t \\ Y_t \end{pmatrix} \sim N\left(\begin{pmatrix} \widehat{X}_{t|t-1} \\ \widehat{Y}_{t|t-1} \end{pmatrix}, \begin{pmatrix} \widehat{\Sigma}_{X_t X_t|t-1} & \widehat{\Sigma}_{X_t Y_t|t-1} \\ \widehat{\Sigma}_{Y_t X_t|t-1} & \widehat{\Sigma}_{Y_t Y_t|t-1} \end{pmatrix}\right)$ to factor out the error variance matrix $\widehat{\Sigma}_{Y_t Y_t|t-1}$ of the observable components. The formulae (5.3.12) and (5.3.13) are the results of the application of Proposition 5.5. Since the estimate of the time-*t* observable components $\widehat{Y}_{t|t-1} = \widehat{Y}_t\left(I_{t-1}\right)$ is linear in I_{t-1}, we have $\widehat{\theta}_s\left(I_t\right)=\widehat{\theta}_s\left(I_{t-1}, y_t\right)=\widehat{\theta}_s\left(I_{t-1}, y_t - \widehat{Y}_{t|t-1}\right)$ by the optimality of the linear least-squares problems. By the orthogonal projection principle of Proposition 5.3, we have both I_{t-1} and $\widehat{Y}_{t|t-1}$ are uncorrelated with the differential vector $y_t - \widehat{Y}_{t|t-1}$. So we can update the estimate $\widehat{\theta}_s\left(I_t\right)$ as $\widehat{\theta}_s\left(I_{t-1}\right)+\widehat{\theta}_s\left(y_t - \widehat{Y}_{t|t-1}\right)$ - μ_s, which results in the recursive formulae (5.3.12) and (5.3.13). ■

The proposition provides the theoretical foundation for our sequential learning and scenario generation algorithm. The scenarios are organized in a tree structure as the following example in Figure 5-3. It serves as a discrete approximation to the parameter process. We have setup some notations to define a scenario tree in §3.2. For convenience, we list a few frequently used notations here.

We only consider symmetric scenario trees whose stage-*t* nodes all have n_t branches. The positive integer n_t is hence the stage-*t* fanning factor. The structure of a T-stage symmetric tree is uniquely determined by fanning factors $\left\{n_0, \cdots, n_{T-1}\right\}$. The tree nodes are indexed by consecutive integer numbers 0, 1, 2, … starting at the root node, from left to right, and from top to bottom, as shown in Figure 5-3. For $t = 0,1,\cdots,T$, the set $\mathbb{N}_t$ denotes all nodes at the stage-*t*. Clearly, $\mathbb{N}_0 = \left\{0\right\}$ only contains the root node, $\mathbb{N}_T$ contains all leaf nodes, and $\mathbb{N} = \underset{t=0,\cdots,T}{\cup} \mathbb{N}_t$ all tree nodes.

To traverse the tree, we define two operators on the set of nodes, a predecessor operator $a\left(\cdot\right): \mathbb{N} \to \mathbb{N}$ which returns the immediate predecessor of the operand and a successor operator $N\left(\cdot\right) \subset \mathbb{N}$ which returns the set of immediate successors of the operand. We use node index s_t to denote a stage-*t* node. For example, for an intermediate stage-*t* node $s_t \in \mathbb{N}_t$ and $1 \le t \le T-1$, it has a unique predecessor

$s_{t-1} = a(s_t)$ and $s_{t-1} \in \mathbb{N}_{t-1}$. Its immediate successors are included in the set $\mathbb{N}(s_t)$, a subset of $\mathbb{N}_{t+1}$ such that $N(s_t) \subseteq \mathbb{N}_{t+1}$ with $a(s_{t+1}) = s_t$, $\forall s_{t+1} \in \mathbb{N}(s_t)$.

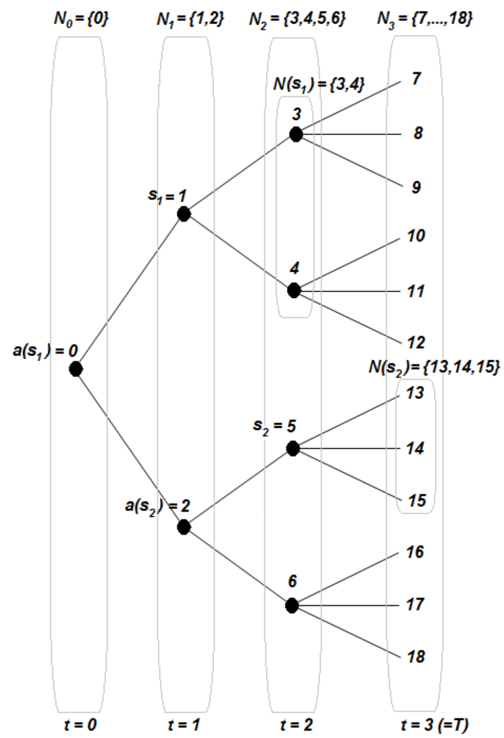


Figure 5-3 A 4-stage symmetric scenario tree with fanning factors $n_0 = 2$, $n_1 = 2$, and $n_2 = 3$. The predecessor operator $a(s_t)$ and the successor operator $\mathbb{N}(s_t)$ are explained.

We use $y_t^{s_t}$ to represent the realized value of $Y_t(s_t)$, the random vector Y_t at node s_t. The transition probability from a stage-*t* node *i* to one of its successor $j \in N(i)$ is $p_{j|i} = \mathbb{P}\{s_{t+1} = j \mid s_t = i\}$ = $\mathbb{P}\{Y_{t+1}(s_{t+1}) = y_{t+1}^j \mid Y_t(s_t) = y_t^i\}$. Each node other than the root is characterized by a value-probability pair $\left(y_t^{s_t},\ p_{s_t|a(s_t)}\right)$. The updated distribution of θ_{t+1} conditional on the node s_t is $\theta_{t+1} \mid y_t^{s_t} \sim N\left(\widehat{\theta}_{t+1|t}^{s_t}, \widehat{\Sigma}_{t+1|t}\right)$.

Scenario generation: a step-by-step example

Before we present the general algorithm for the scenario tree generation, we give a step-by-step description to generate the scenar4io tree in Figure 5-3.

1. Determine the distribution of Y_1 conditional on the root node

Starting at the root node $s_0 = 0$ at stage 0, we don't have any observation to update the prior information $\theta_1 \sim N\left(\mu_1, \Sigma_1\right)$, the parameter distribution at stage *1*. Conditional on the root, the distribution is therefore the same $\theta_1 \sim N\left(\widehat{\theta}_{1|0} = \mu_1, \widehat{\Sigma}_{1|0} = \Sigma_1\right)$. From it, we can extract the distribution $N\left(\widehat{Y}_{1|0}, \widehat{\Sigma}_{Y_1 Y_1|0}\right)$ of the observable components Y_1. In the scenario tree, we use the n_0 successor nodes of the root to approximate the distribution of Y_1. In our case, n_0= 2 and $N\left(0\right) = \left\{1,2\right\}$. Each successor $i \in N\left(0\right)$ is characterized by a value-probability pair $\left(y_1^i, p_{i|0}\right)$, where $y_1^j = Y_1\left(s_1 = j\right)$ and $p_{i|0} = \mathbb{P}\left\{s_1 = i \mid s_0 = 0\right\}$. The next question is how to generate a good discrete approximation $\left\{\left(y_1^i, p_{i|0}\right), i \in N\left(0\right)\right\}$ for the multi-normal distribution $N\left(\widehat{Y}_{1|0}, \widehat{\Sigma}_{Y_1 Y_1|0}\right)$ such that $\left|N\left(0\right)\right| = n_0$.

2. Compute the successor nodes and associated probability

There are different ways to discretize a multivariate continuous distribution, each following a different technique or "goodness of discretization" criterion. In our work, we take a two-step approach which is accurate and efficient. First, we can draw a random sample of size *M* from the multi-normal distribution $N\left(\widehat{Y}_{1|0}, \widehat{\Sigma}_{Y_1 Y_1|0}\right)$ with any Monte Carlo method, where the sample size *M* is much larger than n_0, such as 1000. Second, we apply a scenario reduction algorithm to reduce the large sample to remain only n_0 data points $\left\{y_1^i, i \in N\left(0\right)\right\}$ with reassigned probabilities $\left\{p_{i|0}, i \in N\left(0\right)\right\}$. The ideas and criteria of the scenario reduction algorithm are explained later.

3. Update the prior beliefs with realized values

The successors fanning out from a node simulate the possible states to which the system will transit from the current node (or state). Since the random vector Y_1 is correlated with θ_2, the parameters at stage 2, the realized values of Y_1 provide information to update the distribution of θ_2. So for each node in $N(0)$ at stage 1, we can use the node information to update the belief about the distribution of θ_2 at stage 2. The prior belief is the same $\theta_2 \sim N\left(\widehat{\theta}_{2|0} = \mu_2, \widehat{\Sigma}_{2|0} = \Sigma_2\right)$ for both nodes in $N(0)$. For the realized values $y_1^i = Y_1(s_1 = i)$ at each node $i \in N(0) = \{1,2\}$, given the covariance matrix $\Sigma_{\theta_2 Y_1} = \Sigma_{Y_1 \theta_2}{}'$, we can apply (5.3.12) and (5.3.13) to compute the updated parameters as,

$$\widehat{\theta}_{2|1}^i = \widehat{\theta}_{2|0} + \Sigma_{\theta_2 Y_1} \widehat{\Sigma}_{Y_1 Y_1|0}{}^{-1} \left(y_1^i - \widehat{Y}_{1|0} \right),$$
$$\widehat{\Sigma}_{2|1}^i = \widehat{\Sigma}_{2|0} - \Sigma_{\theta_2 Y_1} \widehat{\Sigma}_{Y_1 Y_1|0}{}^{-1} \Sigma_{Y_1 \theta_2} = \widehat{\Sigma}_{2|1}.$$

From the above updating formulae, we can see the updated covariance matrixes don't depend on the exact realized values of Y_1 and only depend on the distribution of the Y_1. However, the exact values of Y_1 do impact the belief about the point estimates of the parameters. So the updated (multi-normal) distributions at sibling nodes have the same covariance matrix but shifted means.

4. Transit to node 1 and repeat the preceding three steps

We choose the DFS to traverse and perform the simulation-reduction-and-updating procedure to recursively generate the scenario tree. The DFS traversal leads us to transit to node *1* after we finish processing the root as step 1 through step 3 above. At node 1, we have updated belief about the distribution of the parameters at stage 2,

$\theta_2 \sim N\left(\hat{\theta}^1_{2|1}, \hat{\Sigma}_{2|1}\right)$. Like step 2, we apply the simulation-reduction method to generate the n_1=2 successors of node *1*, which are node 3 and node 4, denoted as $N(1)=\{3,4\}$. We can compute the value-probability pairs of nodes in $N(1)$, $\left\{\left(y^i_2, p_{i|1}\right), i \in N(1)\right\}$, which forms a discrete approximation to $N\left(\hat{\theta}^1_{2|1}, \hat{\Sigma}_{2|1}\right)$. Using those (simulated) observations at the nodes in $N(1)$, we can update the belief at each $i \in N(1)$ as $N\left(\hat{\theta}^i_{3|2}, \hat{\Sigma}_{3|2}\right)$.

After we have updated the successor nodes of node 1, we can continue our DFS and transit to node 3. After the transition, we repeat the step 2 again to compute the value-probability pairs of the three successors of node 3, denoted as $N(3)=\{7,8,9\}$. At this time, we have reached the last stage and we need backtrack from node 3 to check if there is any other node on the path back to the root whose successors are not processed yet. The backtracking is part of the DFS.

When we go back from node 3 to its predecessor, node 1, we found the other successor of node 1, node 4, has not been visited. Therefore we transit to node 4 and repeat the same process as we have done at node 3. After we finish the work to visit node 4, we backtrack all the way back to the root and then find the node 2 has not been processed yet. Then we transit to node 2 and repeat all the procedures as we transited to node 1. After we finish the subtree rooted at node 2, we backtrack to the root again. We will found there is no unprocessed successor of the root. So we have generated the scenario tree.

Scenario tree generation: algorithm

The scenario tree includes two parts. The first part is a data structure to represent the topology of the scenario tree. The second part is a list of data associated to the tree nodes. The both parts collaborate to represent the discrete time approximation to our belief of the reservoir performance.

The following gives a general algorithmic description to the scenario tree generation. We choose the DFS-based traversal since the space complexity (required memory) is proportional to the height of the tree, which is usually not too large for practically solvable models. Normally, we choose stack-based data structures to save and retrieve intermediate results about the updated parameters while we dive into the deep end along a path of the tree. We can use a data structure stack, *S*, to keep track of which node will be sampled next.

0. Inputs: $\left(\mu_{\theta_t}, \Sigma_{\theta_t\theta_t}\right)$ and serial covariance matrix $\Sigma_{\theta_s Y_t}$ for all $0 \le t < s \le T$.

1. Initialization:
   a. $\left(\widehat{\theta}_{t|0}, \widehat{\Sigma}_{t|0}\right) = \left(\mu_{\theta_t}, \Sigma_{\theta_t\theta_t}\right)$ and $\left(\widehat{Y}_{t|0}, \widehat{\Sigma}_{Y_tY_t|0}\right) = \left(\mu_{Y_t}, \Sigma_{Y_tY_t}\right)$, for all $t = 1, \ldots T$.
   b. Node index $i = j = 0$, time index $t = 0$, stack $S = \{0\}$.

2. For $t = 0, \ldots T$-1,

   For $\tau = t+1, \cdots, T-1$, we update the covariance matrix with

$$\widehat{\Sigma}_{\tau|t+1} = \widehat{\Sigma}_{\tau|t} - \Sigma_{\theta_\tau Y_t} \widehat{\Sigma}_{Y_tY_t|t}{}^{-1} \Sigma_{Y_t\theta_\tau}; \tag{5.3.14}$$

3. If $S = \{\varnothing\}$, exits and printout "Scenario tree generation is done."

   Otherwise, i = pop(S), t = Stage(i), continue.

4. Draw *M* samples $\{y^m, m = 1, \cdots, M\}$ from the distribution $N\left(\widehat{Y}^i_{t+1|t}, \widehat{\Sigma}_{Y_{t+1}Y_{t+1}|t}\right)$. Call scenario reduction to get the reduced sample $R = \left\{\left(y^{(m)}, p^{(m)}\right), m = 1, \cdots, n_t\right\}$ of

size n_t. Match the elements of the set R to the nodes in $N(i)$ such that $\left\{\left(y_{t+1}^{j}, p_{j|i}\right), j \in N(i)\right\}$ is any permutation of the elements in set *R*.

5. If $t < T-1$,

   1) For each $j \in N(i)$, update

$$\hat{\theta}_{t+2|t+1}^{j} = \hat{\theta}_{t+2|0} + \Sigma_{\theta_{t+2}Y_{t+1}} \hat{\Sigma}_{Y_{t+1}Y_{t+1}|t}^{-1} \left(y_{t+1}^{j} - \hat{Y}_{t+1|t}\right); \qquad (5.3.15)$$

   2) The updated distribution at node *j* is $\theta_{t+2}^{j} \sim N\left(\hat{\theta}_{t+2|t+1}^{j},\ \hat{\Sigma}_{t+2|t+1}\right)$;

   3) For each $j \in N(i)$, push(*S*, *j*).

6. Go to 2.

The relationship to the Kalman filtering

One interesting observation is that our sequential learning algorithm can be treated as a special Kalman filtering with no control involved in the system state transitions. With some changes, we can represent our sequential learning algorithm in the three-step form, prediction-observation-correction, of Kalman filtering.

In Kalman filtering language, we call the parameters θ_t as the system state. We assume the prior information $\left\{\theta_t \sim N\left(\mu_t, \Sigma_t\right), t = 0, \cdots, T-1\right\}$ can be represented as a linear dynamic system,

$$\theta_{t+1} = A_t \theta_t + w_t, \quad t = 0, \cdots, T-1, \qquad (5.3.16)$$

where the initial state $\theta_0 = \mu_0$, the matrixes A_t satisfies $\mu_{t+1} = A_t \mu_t$ and the serial correlation $\Sigma_{\theta_{t+1}\theta_t} = A_t \Sigma_t$, and $w_t \sim N\left(0, M_t\right)$ denotes the model inaccuracy to approximate the prior information. Notice the prediction error of θ_{t+1} recursively accumulates all previous prediction errors according to (5.3.16). If the covariance matrixes $\{M_u, u = 0, \cdots, t\}$ satisfy $\underset{\{w_u, u=0,\cdots,t\}}{\mathbb{E}}\left[(\theta_{t+1}-\mu_{t+1})(\theta_{t+1}-\mu_{t+1})'\right] = \Sigma_{t+1}$, the prior information can be generated by the linear dynamic system (5.3.16).

The Kalman filtering algorithm requires a measurement to reveal all or partial information of the system state θ_t. Since we can observe partial information Y_t of the vector θ_t, the observable components Y_t naturally serves as the measurement to θ_t.

$$Y_t = C_t\theta_t + v_t, \quad t = 0, \cdots, T-1, \tag{5.3.17}$$

where each row of the matrix C_t contains only one nonzero element whose value is one and serves to select a corresponding observable component of θ_t for Y_t. The random disturbance $v_t \sim N(0, N_t)$ captures the measurement errors and the error covariance matrix $N_t = \Sigma_{Y_tY_t}$ can be inferred from the prior information.

The Kalman filtering works in this way. Assume at time *t*, we already have the prediction $\widehat{\theta}_{t|t-1}$ and prediction error covariance matrix $\widehat{\Sigma}_{t|t-1}$ from last iteration.

1. Before the value of Y_t is resolved, we can form a prediction $\widehat{Y}_{t|t-1} = C_t\widehat{\theta}_{t|t-1}$ to it with $\widehat{\theta}_{t|t-1}$ and the state transition equation (5.3.16).
2. During the transition to the next state, the value of Y_t is resolved, which is y_t. Observing this, we can use linear least-squares estimate to correct the previous prediction $\widehat{\theta}_{t|t-1}$ by the following formula,

$$\widehat{\theta}_{t|t} = \widehat{\theta}_{t|t-1} + G_t\left(y_t - \widehat{Y}_{t|t-1}\right). \tag{5.3.18}$$

Notice, both $\hat{\theta}_{t|t-1}$ and $\hat{Y}_{t|t-1}$ are predictions based on previous observations prior to the presence of y_t. When y_t comes, we know the prediction error $y_t - \hat{Y}_{t|t-1}$ of Y_t. Therefore, we can use this knowledge to correct the prediction $\hat{\theta}_{t|t-1}$. The matrix G_t is called the Kalman gain. It works as a weight matrix applied to the correction terms. It is determined such that the mean-square error of the estimate of θ_t defined by the right hand side of (5.3.18) is minimized. The corrected error covariance matrix $\hat{\Sigma}_{t|t}$ can be obtained as well. There are standard methods to compute G_t and $\hat{\Sigma}_{t|t}$. Their formulae can be found on any books on Kalman filtering methods.

3. Once we have the corrected estimate of $\hat{\theta}_{t|t}$, we can predict the next system state $\hat{\theta}_{t+1|t} = A_t \hat{\theta}_{t|t}$ using (5.3.16) and associated error covariance matrix $\hat{\Sigma}_{t+1|t} = A_t \hat{\Sigma}_{t|t} A_t'$. At this time, we enter the step 1 of the next loop.

5.3.3 Advantages of sequential learning

We start the scenario generation with a prior belief about the stochastic process of the parameters $\{\theta_t, t = 1, \cdots, T\}$ with $\theta_t \sim N(\mu_t, \Sigma_t), \forall t$. For any $t \geq 1$, we denote the observation history up to stage t as $I_t = (y_1, y_2, \cdots, y_t)$. After running the sequential learning algorithm above, we end up with an educated process such that the distribution of the parameter θ_{t+1} conditional on any observation history I_t is $\theta_{t+1} \mid I_t \sim N\left(\hat{\theta}_{t+1}(I_t), \hat{\Sigma}_{t+1|t}\right)$, $\forall t$, where $\hat{\theta}_{t+1}(I_t)$ is the linear least-squares estimate of θ_{t+1} given I_t and $\hat{\Sigma}_{t+1|t} = \mathop{\mathbb{E}}_{I_t, \theta_{t+1}}\left[\left(\theta_{t+1} - \hat{\theta}_{t+1}(I_t)\right)\left(\theta_{t+1} - \hat{\theta}_{t+1}(I_t)\right)'\right]$. We can compute both $\hat{\theta}_{t+1}(I_t)$ and $\hat{\Sigma}_{t+1|t}$ recursively by (5.3.15) and (5.3.14). Since the linear least-squares estimate $\hat{\theta}_{t+1}(I_t)$ is an unbiased estimate for θ_{t+1}, we have $\mathop{\mathbb{E}}_{I_t}\left[\hat{\theta}_{t+1}(I_t)\right] = \mu_{t+1}$. Therefore, the unconditional updated distribution of θ_{t+1} is $\hat{\theta}_{t+1} \sim N\left(\mu_{t+1}, \hat{\Sigma}_{t+1|t}\right)$ for all t. We have

$\widehat{\Sigma}_{t+1|t} = \widehat{\Sigma}_{t+1|0} - \Sigma_{\theta_{t+1}Y_t} \widehat{\Sigma}_{Y_tY_t|t-1}{}^{-1} \Sigma_{Y_t\theta_{t+1}}$ by (5.3.14). Since $\widehat{\Sigma}_{Y_tY_t|t-1}$ is positive definite, $\widehat{\Sigma}_{Y_tY_t|t-1}{}^{-1}$ is positive definite as well and the diagonal elements of $\widehat{\Sigma}_{t+1|t}$ are less or equal to the diagonal elements of $\widehat{\Sigma}_{t+1|0} = \Sigma_{t+1}$ component-wisely. In other words, each marginal distribution of the updated (joint) distribution of the parameter θ_t for all $t \geq 1$ has a narrower spread than that before learning. So for the same statistical inference based on the parameter process, the updated process will give a narrower confidence interval to achieve the same confidence level.

5.3.4 Scenario reduction

The scenario reduction is a numerical technique to approximate a discrete distribution of a large sample set with a smaller subset. The basic idea is to assess and minimize a special distance measure between the original distribution and the reduced distribution. There are different measures and methods to achieve good reduction. Hoyland and Wallace (2001), Romisch (2003), and Dupacova et al. (2001) provide extensive reviews and tests.

For simplicity, we use the commercial package, called GAMS/SCENRED, to achieve scenario reduction. The SCENRED is a standard package provided in GAMS for all recent versions since 2002. The detailed mathematics and the user manual please refer to Dupacova et al. (2003) and Heitsch and Romisch (2007, 2009). To make the dissertation a self-contained one, we just provide a brief introduction to the method.

Assume we have two discrete distributions *Q* and *P*, which are characterized by two lists of value-probability pairs $\left\{\left(x^i, q_i\right), i = 1, \cdots, M\right\}$ and $\left\{\left(y^j, p_j\right), j = 1, \cdots, N\right\}$, respectively, where $\sum_i q_i = \sum_j p_j = 1$, $q_i, p_j > 0$, and $x^i, y^j \in \mathbb{R}^d, \forall i, j$. The distance between *P* and *Q* is measured by the Kantorovich functional

$$\mu_c(Q,P) = \inf_{\eta} \left\{ \sum_{i,j} \eta_{ij} c\left(x^i, y^j\right) : \eta_{ij} \ge 0, \sum_j \eta_{ij} = q_i, \sum_i \eta_{ij} = p_j \right\}, \qquad (5.3.19)$$

where η is any discrete distribution measure over all mapping of $R^d \times R^d \to [0,1]$ such that $\eta_{ij} = \mathbb{P}_{\eta}\left(X = x^i, Y = y^j\right)$ and $c(\cdot): R^d \times R^d \to R_+$ is a probability metric function such that the underlying stochastic optimization model is stable with respect to μ_c. Normally, we choose the metric function $c(\cdot)$ such that it is continuous and symmetric and serves as an atomic distance function such that $c\left(x^i, y^j\right) \ge 0$ with "=" being true iff $x^i = y^j$. For example, one typical choice is $c(x,y) = \|x-y\|_{\infty}$.

The problem (5.3.19) is called Monge-Kantorovich mass transportation problem. It aims to find an optimal distribution measure η^* such that η^*-weighted probability distance between Q and P is minimal under the measure $c(\cdot)$. Notice (5.3.19) is a linear program. Then we can write down its dual form as,

$$\mu_c(Q,P) = \sup_{u,v} \left\{ \sum_i q_i u_i + \sum_j p_j v_j : u_i + v_j \le c\left(x^i, y^j\right), \forall i, j \right\}. \qquad (5.3.20)$$

The problem to perform scenario reduction can be interpreted as follows. Given a discrete distribution Q with a large number of scenarios $\left\{\left(x^i, q_i\right), i = 1, \cdots, M\right\}$, we want to find the best approximation of Q with respect to μ_c by such a discrete distribution P consisting of scenarios $\left\{\left(y^j, p_j\right), j = 1, \cdots, N\right\}$ such that $\left\{y^j\right\}_{j=1}^N$ is a subset of $\left\{x^i\right\}_{i=1}^M$ and $N << M$. Clearly, this is a combinatorial optimization problem and normally solved with efficient heuristic methods based on both the primal and dual forms.

The details about the forward algorithm and backward algorithm to solve such problems can be found in the papers mentioned above. The manual to use the

GAMS/SCENRED can be found among the official help files on any installed GAMS system or downloaded from the official website of GAMS. The two-volume book by Rachev and Ruschendorf (1998) provides extensive discussions in theory and applications of the Monge-Kantorovich mass transportation problem.

5.4 NUMERICAL EXPERIMENTS

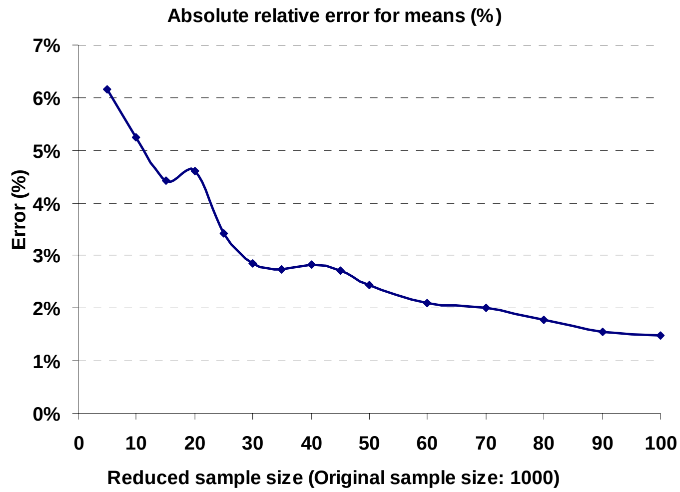
We implement the scenario generator in GAMS. The generator consists of two modules. The first module receives user's inputs, which include the solution quality requirements (measured by fanning factors) and assessments about the means and correlations of parameter for all oil projects. For test purposes, we assume hypothetical projects and draw samples from some reasonable ranges to simulate their geological profiles. The second module takes the inputs to generate the two subtrees. We have presented methods to generate the two subtrees in §5.2 and §5.3, respectively.

Besides the methods introduced in previous sections, we need two additional numerical techniques to implement the generator. The first technique is to generate random samples from a multinormal distribution with known mean and variance-covariance matrix. Gentle (2003) and Glasserman (2004) provides reviews to some frequently used methods to generate multinormal samples. We choose the linear transformation approach, which multiplies the Cholesky factor of the variance-covariance matrix to a vector of independent standard normal variables and shifts the product by the mean vector. The second technique is required to ensure the consistence between the correlations implied by the original data and by the logged data. The insurance is discussed in the Appendix of Wang (1998). The key idea is to properly calibrate the mean vector and covariance matrix of the multinormal distributions of the logged data such that when we take exponential to the sampled data from the multinormal distributions, the resulting sample's correlation structure is the same as the original data. With those techniques, we can generate a scenario tree with previous methods.

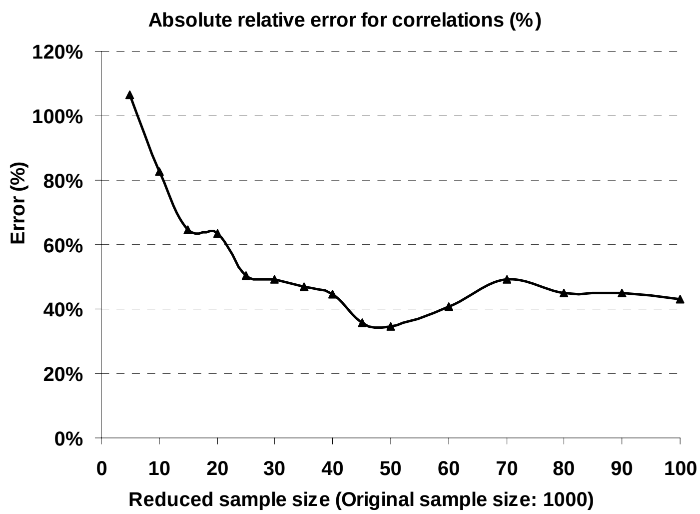
Since the scenario tree approximates the evolution of a random vector while statistical learning is taken into account, we are interested in the goodness of the approximation and learning behavior. Particularly, the ultimate scenarios are selected and probability-adjusted by the scenario reduction code from a large equal-probability random sample generated by Monte Carlo simulation. Therefore, we also want to know how good the reduction code is.

To answer those questions, we run numerical experiments to study the approximation quality. We choose the (averaged) absolute relative error (ARE) as the quality measure and the inputs as the comparison basis. As an example, we only pay attention to the first phase scenarios which approximate two parameters of the tank model, resv and ip. We generate 1000 sample points using the linear transform method and a standard Monte Carlo method. Our test shows the ARE measure for sample means is only 0.8%, which is accurate enough. However, the ARE measure for sample correlations is much worse and around 14%. So compared to matching the mean vector, matching the correlation matrix is a more challenging task.

Since scenario reduction selects a much smaller subset of samples from the above sample with adjusted probabilities, the reduced samples form a cruder approximation to the original distribution. The following Figure 5-4 (a) and (b) show how the approximation quality varies as we change the reduced sample size.



(a)



(b)

Figure 5-4 The approximation quality of the reduced samples. (a) measures the approximation quality for the reduced sample's means and (b) measures the quality for reduced sample's correlations.

The general trends for both curves show that the approximation quality (measured by the AREs of sample means and correlations) can be improved if we increase the size of the reduced sample. The improvement is not necessary monotonic since all data are based on one realization of Monte Carlo simulations. We can achieve good approximation to means with relatively small samples. However, the approximation to correlations is not satisfactory. Even when we increase the reduced sample size to 100, nearly half of the correlations are distorted. So how to find reduced samples which not only lead to good optimal solutions but also achieve accurate statistical approximation could be a continuing research topic.

Chapter 6 Conclusions and Future Work

This chapter concludes the dissertation with a summary of major results and a to-do list for future research topics. The details are arranged as follows. In §6.1, we summarize the major results and conclusions in the dissertation, chapter by chapter. In §6.2, we list a few things to do as future work which would further consummate the MEPPO model in both modeling and solution efficiency.

6.1 CONCLUSIONS

This dissertation has developed multistage stochastic optimization models to assess and prioritize multiple oil fields and manage their risk in a portfolio approach. To quantify the impacts of the market risk and the field-specific risk, we develop both real options-based and decision tree-based optimization valuation models to capture respective risks and field development decisions. By pooling cash flows incurred by the decisions of selected oil fields, we enforce dynamic budget constraints and hence create a portfolio optimization model to screen and develop oil fields. The optimization goal is to maximize some mean-risk multi-criteria objective function. The solution gives the optimal contingent plans to develop oil fields in a portfolio perspective. The advantages for the portfolio approach to manage multiple oil fields are manifold, including risk hedging, synergy effects, and consistence.

In the first two chapters, we introduce the background knowledge of exploration and production business and provide a literature review to oil project valuation methods and risk management. The introduction serves as the foundation to develop a concise phase model to capture major decisions and information for typical E&P projects. At the end of Chapter 2, we generalize and extend the classic static and nonlinear tank model to a dynamic and linear one with MIP techniques. The generalized model is used in Chapter 3 to forecast reservoir production performance as sequential drilling decisions are allowed.

In Chapter 3, we consider the portfolio optimization model for explored oil fields whose geologic risk has been largely reduced by earlier explorations. As a consequence, the market risk (due to oil price volatility) plays a dominating role to determine the fields' development strategies and hence their values. By assuming a complete market, we can model and evaluate each oil field as a sequential compound option. In this sense, our project portfolio model is essentially a portfolio of real options. For this purpose, we develop two optimization frameworks to price financial or real options. The first framework is based on the classic method to solve infinite horizon stochastic dynamic programming (SDP) problems with linear programs. We extend the method to deal with both infinite and finite horizon SDP problems with linear programs, which allow us to price both perpetual options and options with finite expiries with linear programming. The second framework is to formulate the optimal exercising (or stopping) problem to price an option as a multistage stochastic integer programming problem. In this framework, the exercise decisions are explicitly captured by binary decision variables.

Both frameworks rely on a state tree (or scenario tree) to capture the underlying price dynamics. If the price follows a GBM process, we can conveniently use the binomial lattice or trinomial lattice as the state tree. If the price follows a mean-reverting process, we can use a drifted trinomial lattice. For other price processes, we can use general scenario tree generation approaches to create the scenario tree.

The LP based pricing model has a higher computational performance while the MIP model endows us more flexibility to deal with more complex decisions. It is more convenient and straightforward to use the MIP option pricing model to model the portfolio of real options since the drilling decisions and other investment decisions can be explicitly attacked. As an example, we apply the MIP option pricing framework to model the oil field development problem which allows drilling decisions to be made sequentially. We further present a real options based dynamic capital budgeting model to screen and develop multiple oil fields using a limited initial budget. The numerical experiments show the effectiveness and advantages of the portfolio approach to maximize the overall benefits of multiple real options. However, since the MIP option pricing model involves a large number of discrete variables, how to extend its applications to deal with more practical problems poses us a new challenge. One likely resolution is to integrate both the LP pricing model and the MIP pricing model and notice the constraints of the LP model can be generated in a cut-generation approach. We would leave this challenge as one topic for our future research.

In Chapter 4 and 5, we turn to develop the multistage E&P project portfolio optimization (MEPPO) model. The MEPPO model consists of three major components, a scenario tree expressing the evolution of all random parameters, a number of decision trees modeling individual E&P projects, and a sequence of budget constraints mandating all projects to share the same cash account.

We first present a phase model to organize main E&P decisions and associated information structures on which the decisions are made. Since the phase model covers the full life cycle of an E&P project, it should consider only the most important elements. Based on the phase model, we model the sequential investment process of an E&P project as a decision tree and formulate the decision tree as an optimization model with

MIP techniques. This approach allows us to directly capture the cash flows incurred to a project. Assuming a common state tree (scenario tree) which carries at each state node the jointly revealed information of all E&P projects, we can aggregate the cash flows cross projects to form a budget constraint in each state at each stage. The budget constraints involve an inventory variable to carry cash excess from one period to the succeeding period as part of the budget for that period. Our MEPPO model aims to trade off the overall rewards, i.e., the ENPV of the portfolio, against the expected downside deviation.

Given a scenario tree, the MEPPO model can be equivalently formulated as a deterministic large scale MIP. To solve the MIP, we develop an optimization based heuristic method, called DCMP, which breaks the decision process into two parts and further decomposes the second part into a number of subproblems. The first part involves the capital expenditure decisions, such as acquisition, exploration, capacity installation, and so on. These investment activities resolve most geologic uncertainties prior to the production phase. The second part addresses the drilling decisions.

The DCMP method works as follows. It first solves a problem called RMIP, which is essentially the MEPPO problem with integral drilling decisions being relaxed. The DCMP method fixes the first part decisions of the MEPPO problem according to the RMIP solution. The restricted MEPPO problem therefore is decoupled into a number of deterministic drilling scheduling subproblems, one for each sample path of the scenario tree. The subproblems are independent to each other and can be solved very efficiently. After we have solved all drilling subproblems, we further fix the drilling decisions of the MEPPO problem according to the optimal solutions of the subproblems. So far, all discrete variables of the MEPPO problem have been fixed. The DCMP method always ensures the feasibility of the restricted MEPPO problem since the trivial solution (all discrete variables being zero) is always feasible. The restricted MEPPO problem can be solved fast to recover the cash flows corresponding to the fixed decisions. Our numerical experiments have shown the DCMP method is very promising. Since the RMIP is one of the most appropriate and straightforward relaxations of the MEPPO problem, it is usually required to solve to provide an upper bound. We can treat the solution time of the RMIP as an overhead. Once the RMIP is solved, the DCMP method runs fast and its performance is very stable. Based on the extensive tests, the optimality gap of the integer DCMP solution to the MEPPO problem is empirically less than 5%.

Since E&P projects are long-term learning investments, the MEPPO model relies on a probabilistic model to describe the dynamics of multiple E&P projects as an outcome of sequential investments and information updating. The resulting stochastic process for the parameters of the E&P projects plays an indispensable role to find the best project mix and look for the optimal contingent plans to develop the projects.

We use the entire Chapter 5 to discuss how to model the investor's beliefs about the multivariate stochastic process and how to approximate it with a scenario tree. Particularly, we have incorporated different statistical learning methods to update the prior beliefs as investment activities progress. The learning is supported by statistical dependences among the projects, which can be measured by joint probabilities, conditional probabilities, or correlations. For E&P projects, we identify two types of dependences, the intra-project dependence and the inter-project dependence. The former takes a multivariate binary distribution form to describe the joint state of "pay/no pay" of all oil fields. The latter uses a multivariate continuous distribution to represent how the tank model's parameters of each field co-move over time. Correspondingly, we respectively apply the optimal binary learning method and the sequential linear least-squares estimation method (or sequential Bayesian updating as a special case) to integrate the information contained in the dependences to update the priors. We further show that the sequential updating procedure ensures that the educated process of the parameters asymptotically converges to the population means.

Once we have setup the probabilistic models, we can generate a scenario tree to approximate the dynamics of the E&P project portfolio. Since there are two types of dependences of disparate natures, we first separately generate two subtrees and then combing them to form a composite scenario tree by taking Cartesian product over the state paces of the two subtrees. The first subtree simply models the joint binary distribution. The second subtree is generated recursively with Monte Carlo simulation methods.

It is known that the quality of the Monte Carlo simulation (and its variants) relies on the Law of Large Number (LLN) theorem. Monte Carlo simulation requires a sufficiently large sample to guarantee the goodness of approximation (to the original distribution). So a naive application of the simulation to generate the scenario tree will unavoidably lead to a huge tree with hundreds of branches fanning out of any node. What makes this even worse is that the state space of the scenario tree is of high dimension. The remedy we provide to alleviate the curse of dimensionality is to perform scenario reduction to the branches out of each node. The idea is to use a limited number of branches with adjusted probabilities to substitute the originally generated equal-likely branches. So at each node, we first generate a large number of sample points with Monte Carlo simulation techniques and then we perform reduction to the large sample to only leave a few sample points. The probabilities on those sample points are determined to match the moments of the original sample. Again, the moment matching is multivariate.

One major advantage of this simulation-reduction approach is that the growth rate and the goodness of approximation of the scenario tree are always under our control, no matter how many projects will be considered and how many parameters will be involved to forecast reservoir production. We do trade off the accuracy against solvability via the scenario reduction. However, the price we pay is worth of the value we gain. Our extensive numerical experiments have verified the quality and efficiency of this approach.

6.2 FUTURE WORK

The MEPPO model is a comprehensive framework. It covers various decisions and information stages over the entire life-cycle of typical E&P projects. We have to sacrifice the completeness and make a sequence of simplifications in order to get a solvable model. One major simplification is to ignore the market risk and assume a fixed oil price (or a fixed forward curve for oil prices) to estimate future revenues. So the MEPPO model ignores trading opportunities. A natural extension would be to incorporate trading opportunities and consider the market risk.

Even though the MEPPO model has been largely simplified, the resulting MIP problem is still very difficult to solve, which is especially true when the number of projects increases. The DCMP method is a promising approach to solve large instances. However, it relies on a good solution of the RMIP problem, which itself is difficult to solve. Therefore, we need expedite the solution of the RMIP problem without sacrificing the solution quality.

To conclude this dissertation, we discuss the above two directions of future research to make the MEPPO more practical in the following two subsections.

6.2.1 The acceleration of the RMIP solution

The DCMP method decomposes a large scale MIP problem into a number of drilling scheduling subproblems, each of which can be solved very fast. The method itself is very efficient and robust. However, it relies on a good solution of the RMIP problem. When the quality of the RMIP solution is high, the optimality gap of the DCMP solution is empirically around 3% and usually less than 5%. However, the RMIP problem can become very difficult to solve when the MEPPO problem becomes very large. In that case, the RMIP problem may take a long time to run and only returns a poor solution, which makes the quality of the DCMP solution deteriorate as well. The optimality gap of the DCMP solution can rise beyond 30%. Therefore, we need better ways to solve the RMIP problem to guarantee the availability of high quality solutions.

Since the DCMP problem is still a large MIP, one quick thought is to find better bounds to accelerate the branch-and-bound procedure. One idea is to solve the Lagrangian dual problem by dualizing the budget constraints before the production phase begins, i.e., prior to stage t_3 in Figure 4-3. The resulting Lagrangian subproblem can be decoupled into a number of deterministic E&P project portfolio optimization problems, one for each sample path. It is possible to develop some prmal heuristic method based on the Lagrangian relaxation method.

6.2.2 The ideal project portfolio model

This dissertation has been motivated by two streams of independent thoughts on the valuation of individual risky projects and the risk management of multiple projects in a portfolio approach. The first thought is to integrate real options and decision analysis to assess individual E&P projects so that both the project-specific risk (local risk) and the market risk (global risk) are simultaneously taken into account in the valuation. This idea is introduced by Smith and Mccardle (1998). The second thought is to manage project risk in a portfolio scope, which is proposed by Ball and Savage (1999) and Gustafsson and Salo (2006). One would be interested in synthesizing both thoughts to get an ideal project portfolio optimization model which simultaneously applies the option pricing theory to hedge the global risk and the decision tree analysis to manage the local risk in a holistic portfolio view.

This dissertation reflects some efforts toward this goal. We have developed two dynamic portfolio optimization models for oil projects of two different types. One model creates a portfolio of real options to deal with explored oil fields in Chapter 3. The other model constructs a portfolio of decision trees to handle E&P projects in Chapter 4 and 5. In this sense, we have "partially" achieved the goal to get the proposed unified framework to manage both the market risk and the project-specific risk in the project portfolio.

The "missing" part is the integration of both portfolio models into a synthesized one. In principle, we can close the gap conveniently with our optimization models. In practice, however, the integration of both the global risk and the local uncertainties raises a challenging task to us purely in computation.

References

Azelton, A.M. and A.S. Teufel. 2009. Fisher Investments on Energy. John Wiley $ Sons, Inc., Hoboken, N.J.

Babayev, D.A. 1975. Mathematical Models for Optimal Timing of Drilling on Multilayer Oil and Gas Fields. Management Science, Vol. 21, No. 12, pp. 1361-1369.

Babusiaux, D., S. Barreau, P.R. Bauquis, N. Bret-Rouzaut, A. Chétrit, P. Copinschi, J.P. Favennec, R. Festor, E. Feuillet-Midrier, M. Grossin, D. Guirauden, V. Lepez, and M. Valette. 2004. Oil and Gas Exploration and Production: Reserves, Costs, Contracts. Editions TECHNIP, Paris.

Ball, B.C. B. and S.L. Savage. 1999. Notes on Exploration and Production Portfolio Optimization.

Benkherouf, L. 1994. A Generalized Oil Exploration Problem. European Journal of Operational Research, Vol. 73, pp. 423-429.

Berger, J.O. 1985. Statistical Decision Theory and Bayesian Analysis. 2nd ed. Springer-Verlag, New York Inc.

Bertsekas, D.P. 2007. Dynamic Programming and Optimal Control, Volume I and II. 3rd ed. Athena Scientific.

Bickel, J.E. and J.E. Smith. 2006. Optimal Sequential Exploration: A Binary Learning Model. Decision Analysis, Vol. 3, No. 1, pp. 16–32.

Biller, B. and S. Ghosh. 2006. Chapter 5: Multivariate Input Processes. Handbook in Operations Research and Management Science, Vol.13, Elsevier B.V.

Birge, J.R. and F. Louveaux. 1997. Introduction to stochastic programming. Springer, New York.

Brealey, R.A., S.C. Myers, and A.J. Marcus. 2008. Fundamentals of Corporate Finance. McGraw-Hill, Irwin.

Broadie, M. and J.B. Detemple. 2004. Option Pricing: Valuation Models and Applications. Management Science, Vol. 50, No. 9, pp. 1145-1177.

Brosch, R. 2008. Portfolios of Real Options. Springer-Verlag, Berlin, Heidelberg.

Cho, M.J. and R.H. Stockbridge. 2002. Linear Programming Formulation for Optimal Stopping Problems. SIAM Journal of Control and Optimization, Vol. 40, No. 6, pp.1965-1982.

Christensen, S. 2011. A method for pricing American options using infinite linear programming. Submitted to Computational Finance.

Clemen, R.T. 1986. Calibration and Aggregation of Probabilities. Management Science, Vol. 32, No. 3, pp. 312-314.

Clemen, R.T. 1989. Combining Forecasts: A Review and Annotated Bibliography. Internat. J. Forecasting, 5, P.559–583.

Clemen, R.T. and T. Reilly. 1999. Correlations and Copulas for Decision and Risk Analysis. Management Science, Vol. 45, No. 2, pp. 208-224.

Clemen, R.T. and T. Reilly. 2004. Making Hard Decisions with DecisionTools Suite (2nd Rev. Ed.).South-Western College Pub.

Clemen, R.T., and R.L. Winkler. 1999. Combining Probability Distributions from Experts in Risk Analysis. Risk Analysis, 19, pp. 187-203.

Clemen, R.T., G.W. Fischer, and R.L. Winkler. 2000. Assessing Dependence: Some Experimental Results. Management Science, Vol. 46, No. 8, pp.1100–1115.

Clewlow, L. and C. Strickland. 1998. Implementing derivatives models. John Wiley & Sons Ltd.

Consigli, G., M.A.H. Dempster. 1998. Dynamic Stochastic Programming for Asset-Liability Management. Annals of Operations Research, Vol. 81, pp. 131-162.

Copeland, T. and V. Antikarov. 2003. Real Options: A Practitioner's Guide. Revised Edition. Thomson Texere.

Cottle, R.W. and R.F. Veinott. 1972. Polyhedral Sets Having a Least Element. Mathematical Programming, Vol. 3, pp. 238-249.

Cottle, R.W. and J.S. Pang. 1978. On Solving Linear Complementarity Problems as Linear Programs. Mathematical Programming Study, Vol. 7, pp. 88-107.

Cozzolino, J.M. 1974. Portfolio of Risky Projects. Decision Sciences, Vol. 5, Issue 4, pp.575-586.

Cozzolino, J.M. 1977. Management of Oil and Gas Exploration Risk. Cozzolino Associates, Inc., West Berlin, N.J.

Cryer, C.W. and M.A.H. Dempster. 1980. Equivalence of Linear Complementarity Problems and Linear Programs in Vector Lattice Hilbert Spaces. SIAM J. Control and Optimization, Vol. 18, No. 1, pp.76-90.

de Farias, D. and B. Van Roy. 2003. The Linear Programming Approach to Approximate Dynamic Programming. Operations Research, Vol. 51, No. 6, pp. 850-865.

Dempster, M.A.H. and D.G. Richards. 2000. Pricing American Stock Options Fitting the Smile. Mathematical Finance, Vol. 10, No. 2, pp. 157-177.

Dempster, M.A.H. and J.P. Hutton. 1997. Fast Numerical Valuation of American, Exotic, and Complex Options. Applied Mathematical Finance, Vol. 4, pp. 1-20.

Dempster, M.A.H. and J.P. Hutton. 1999. Pricing American Stock Options by Linear Programming. Mathematical Finance, Vol. 9, No. 3, pp. 229-254.

Dempster, M.A.H., J.P. Hutton, and D.G. Richards. 1998. LP valuation of exotic American options exploiting structure. Journal of Computational Finance, Vol. 2, No. 1, pp. 61-84.

Devine, M. and W. Lesso. 1972. Models for the Minimum Cost Development of Offshore Oil Fields. Management Science, Vol. 18, No. 8(B), pp. 378-387.

Dias, M.A.G. 2004. Valuation of Exploration and Production Assets: an Overview of Real Options Models. Journal of Petroleum Science and Engineering, Vol. 44, pp. 93-114.

Dixit, A.K. and R.S. Pindyck. 1994. Investment under Uncertainty. Princeton University Press, Princeton, N.J.

Dupačová, J., G. Consigli, and S.W. Wallace. 2000. Scenarios for multistage stochastic programs. Annals of Operations Research, Vol. 100, pp. 25-53.

Dupačová, J., N. Gröwe-Kuska, and W. Römisch. 2003. Scenario reduction in stochastic programming: An approach using probability metrics. Mathematical Programming, Vol. 95, No. 3, pp. 493-511.

Flam, S.D. 2008. Option Pricing by Mathematical Programming. Optimization, Vol. 57, No. 1, pp. 165-182.

Frair, L. and M. Devine. 1975. Economic Optimization of Offshore Petroleum Development. Management Science, Vol. 21, No. 12, pp. 1370-1379.

Gentle, James E. 2003. Random Number Generation and Monte Carlo Methods. 2nd ed. Springer.

Ghasemzadeh, F., N. Archer, and P. Iyogun. 1999. A Zero-one Model for Project Portfolio Selection and Scheduling. Journal of the Operational Research Society, Vol. 50, pp. 745-755.

Glasserman, P. 2004. Monte Carlo Methods in Financial Engineering. Springer-Verlag, New York.

Gondzio, J. and R. Kouwenberg. 2001. High-Performance Computing for Asset-Liability Management. Operations Research. Vol. 49, No. 6, pp. 879-891.

Grayson, C.J. 1960. Decisions under Uncertainty, Drilling Decisions by Oil and Gas Operators. Harvard Business School, Division of Research, Boston, MA.

Gustafsson, J. and A. Salo, 2005. Contingent Portfolio Programming for the Management of Risky Projects. Operations Research, Vol. 53, , No. 6, pp. 946-956.

Haugland, D., A. Hallefjord, and H. Asheim. 1988. Models for Petroleum Field Exploitation. European Journal of Operational Research, 37, pp. 58-72.

Heidenberger, K. 1996. Dynamic Project Selection and Funding under Risk: A Decision Tree Based MILP Approach. European Journal of Operational Research, 95, pp. 284-298.

Heikkinen, V. 2003. Timber Harvesting as a Part of the Portfolio Management: A Multiperiod Stochastic Optimization Approach. Management Science, Vol. 49, No. 1, pp. 131-142.

Heitsch, H. and W. Romisch. 2003. Scenario Reduction Algorithms in Stochastic Programming. Computational Optimization and Applications, 24, pp. 187-206.

Heitsch, H. and W. Römisch. 2007. A Note on Scenario Reduction for Two-Stage Stochastic Programs. Operations Research Letters, 35, pp. 731–738.

Heitsch, H. and W. Römisch. 2009. Scenario tree modeling for multistage stochastic programs. Mathematical Programming, Vol. 118, No. 2, pp. 371-406.

Helmes, K. 2002a. Pricing Perpetual Russian Options using Linear Programming, Proceedings of the Fourth Asian Control Conference, pp. 1557-1562.

Helmes, K. 2002b. Numerical Methods for Optimal Stopping Using Linear and Non-linear Programming. Lecture Notes in Control and Information Sciences, Vol. 280, Stochastic Theory and Control, pp. 185-203, Springer-Verlag.

Helmes, K., S. Rohl, and R.H. Stockbridge. 2001. Computing Moments of the Exit Time Distribution for Markov Processes by Linear Programming. Operations Research, Vol 49, No. 4, pp. 516-530.

Hoyland, K. and S.W. Wallace. 2001. Generating Scenario Trees for Multistage Decision Problems. Management Science, Vol. 47, No. 2, pp. 295-307.

Huang, C.F. and R.H. Litzenberger. 1988. Foundations for Financial Economics. North-Holland.

Hultzsch, P.A. 2005. Decision and Risk Analysis through the Life of the Field. Master Thesis, The University Of Texas At Austin.

Hyne, N.J. 2001. Nontechnical Guide to Petroleum Geology, Exploration, Drilling, and Production. 2nd ed. Pennwell Corp.

Jonsbraten, T.W. 1998. Oil Field Optimization under Price Uncertainty. Journal of the Operational Research Society, 49, pp. 811-818.

Jornsten, K.O. 1992. Sequencing Offshore Oil and Gas Fields under Uncertainty. European Journal of Operational Research, 58, pp. 191-201.

Kalman, R.E. 1960. A New Approach to Filtering and Prediction Problems. Transaction of the ASME Journal of Basic Engineering, 82 (Series D), pp. 35-45.

Kaufman, G.M. 1963. Statistical Decision and Related Techniques in Oil and Gas Exploration. Prentice-Hall Inc., Englewood Cliffs, N.J.

Keefer, D.L. 1994. Certainty Equivalents for Three-point Discrete-distribution Approximations. Management Science, Vol. 40, No. 6, pp. 760-773.

Keeney, R.L. and H. Raiffa. 1976. Decisions With Mutiple Objectives: Preferences and Value Tradeoffs. John Wiley and Sons, New York.

King, A. 2002. Duality and Martingales: A Stochastic Programming Perspective on Contingent Claims. Mathematical Programming, Series B, Vol. 91, pp. 543-562.

Klaassen, P. 1998. Financial Asset-Pricing Theory and Stochastic Programming Models for Asset/Liability Management: A Synthesis. Management Science, Vol. 44, No. 1, pp. 31-48

Korn, R. and E. Korn. 2000. Option Pricing and Portfolio Optimization: Modern Methods of Financial Mathematics. American Mathematical Society.

Kouwenberg, R. 2001. Scenario Generation and Stochastic Programming Models for Asset Liability Management. European Journal of Operational Research, 134, pp. 279-292.

Lasdon, L.S., Coffmann, Jr., P.E., Macdonald, R., Mcfarland, J.W., and Sepehrnoori, K. 1996. Optimal Hydrocarbon Reservoir Production Policies. Operations Research, Vol. 34, No. 1, pp. 40-54.

Leffler, W.L., R. Pattarozzi, and G. Sterling. 2003. Deepwater Petroleum Exploration & Production: A Nontechnical Guide. PennWell Corporation.

Longstaff, F.A. and E.S. Schwartz. 2001. Valuing American Options by Simulation: A Simple Least-Squares Approach. The Review of Financial Studies, Spring 2001, Vol. 14, No. 1, pp. 113-147.

Luehrman, T.A. 1998. Strategy as a Portfolio of Real Options. Harvard Business Review. Sep-Oct, 1998, pp.89-99.

Luenberger, D.G. 1984. Linear And Nonlinear Programming. 2nd ed. Springer.

Luenberger, D.G. 1998. Investment Science. Oxford University Press.

Lund, M.W. 2001. Valuing Flexibility in Offshore Petroleum Projects. Annals of Operations Research, 99, pp. 325-349.

Lyons, W.C. 1996. Standard Handbook of Petroleum and Natural Gas Engineering. Gulf Professional Publishing, Houston, TX.

Makridakis, S and R.L. Winkler. 1983. Averages Of Forecasts: Some Empirical Results. Management Science, Vol. 29, No. 9, pp. 987-996.

Manne, A.S. 1960. Linear Programming and Sequential Decision. Management Science, Vol. 6 No. 3, pp. 259-267.

Mas-Colell, A., M.D.Whinston, and J. R. Green. 1995. Microeconomic Theory. Oxford University Press.

McCray, A.W. 1975. Petroleum Evaluations and Economic Decisions. Prentice-Hall, Inc., Englewood Cliffs, N.J.

Mcfarland, J.W., L. Lasdon, and V. Loose. 1984. Development Planning and Management of Petroleum Reservoir Using Tank Models and Nonlinear Programming. Operations Research, Vol. 32, No. 2, pp. 270-289.

Meier, H., N. Christofides, and G. Salkin. 2001. Capital Budgeting under Uncertainty - An Integrated Approach Using Contingent Claim Analysis and Integer Programming. Operations Research, Vol. 49, No. 2, pp. 196-206.

Meister, B., J.M.C. Clark, and N. Shah. 1996. Optimisation of Oilfield Exploration under Uncertainty. Computers and Chemical Engineering, Vol. 20, Suppl., pp. S1251-S1256.

Merrick, J.R.W., J.R. van Dorp, and A. Singh. 2005. Analysis of Correlated Expert Judgements from Extended Pairwise Comparisons. Decision Analysis, Vol. 2, No. 1, pp. 17-29.

Miller, D.J. and W. Liu. 2002. On the Recovery of Joint Distributions from Limited Information. Journal of Econometrics, 107, pp. 259-274.

Newendorp, P.D. and J. Schuyler. 2000. Decision Analysis for Petroleum Exploration. 2nd ed. Planning Press, Aurora, CO.

Orban, J. 1997. Money In The Ground: Insider's Guide to Oil and Gas Deals. 4th ed. Meridian Press, Oklahoma City.

Pflug, G.Ch. 2000. Scenario Tree Generation for Multiperiod Financial Optimization by Optimal Discretization. Mathematical Programming, Vol. 89, No. 2, pp. 251-271.

Pinar, M.C. and A. Camci. 2009. Pricing American Contingent Claims by Stochastic Linear Programming. Optimization, Vol. 58, No. 6, pp. 627-640.

Pinar, M.C. and A. Camci. 2010. An Integer Programming Model for Pricing American Contingent Claims under Transaction Costs. Computational Economics, forthcoming, Online First.

Rachev, S.T. and L. Ruschendorf. 1998. Mass Transportation Problems (Volume I: Theory, Volume II: Applications). Springer.

Ritchken, P.H. 1985. On Option Pricing Bounds. Journal of Finance, Vol. XL, No. 4, pp. 1219-1233.

Rockafellar, R.T. and R.J. Wets. 1991. Scenarios and Policy Aggregation in Optimization under Uncertainty. Mathematics of Operations Research, Vol. 16, Issue 1, pp. 119-147.

Romisch, W. 2003. Approximations of Stochastic Programs: Scenario Tree Reduction and Construction. GAMS Workshop, Heidelberg, Sept. 1-3, 2003. (http://www.gams.com)

Ruszczynski, A. 1997. Decomposition Methods in Stochastic Programming. Mathematical Programming, 79, pp. 333-353.

Schweitzer, P.J. and A. Seidmann. 1985. Generalized Polynomial Approximations in Markovian Decision Processes. Journal of Mathematical Analysis and Applications, Vol. 110, No. 2, pp. 568-582.

Sen, S., L. Yu, and T. Genc. 2006. A Stochastic Programming Approach to Power Portfolio Optimization. Operations Research, Vol. 54, No. 1, pp. 55-72.

Shapiro, A. 2006. On Complexity of Multistage Stochastic Programs. Operations Research Letters, 34, pp. 1–8.

Shreve, S.E. 2004. Stochastic Calculus for Finance II: Continuous-Time Models. Springer Science + Business Media, Inc.

Smit, H.T.J. and L. Trigeorgis. 2004. Strategic Investment: Real Options and Games. Princeton University Press, Princeton, N.J.

Smith, J.E. 1993. Moment Methods for Decision Analysis. Management Science. Vol. 39, No. 3, pp. 340-358.

Smith, J.E. and K.F. Mccardle. 1998. Valuing Oil Properties: Integrating Option Pricing and Decision Analysis Approaches. Operations Research, Vol. 46, No. 2, pp.198-217.

Smith, J.E. and K.F. Mccardle. 1999. Options in the Real World: Lessons Learned in Evaluating Oil and Gas Investments. Operations Research, Vol. 47, No. 1, pp.1-15.

Smith, J.E. and R.F. Nau. 1995. Valuing Risky Projects: Option Pricing Theory and Decision Analysis. Management Science. Vol. 41, No.5, pp. 795-816.

Stockbridge, R.H. 1990. Time-Average Control of Martingale Problems: A Linear Programming Formulation. Annals of Probability, Vol. 18, No. 1, pp. 206-217.

Stockbridge, R.H. 2004. Characterizing Option Prices by Linear Programs. Mathematics of Finance, G. Yin and Q. Zhang eds., AMS Contemporary Mathematics, Vol.351, pp. 349-359.

Suslick, S.B. and D.J. Schiozer. 2004. Risk Analysis Applied to Petroleum Exploration and Production: an Overview. Journal of Petroleum Science and Engineering, Vol. 44, pp. 1-9.

Trigeorgis, L. 1996. Real Options: Managerial Flexibility and Strategy in Resource Allocation. Mit Press, Cambridge, MA.

Vanderbei, R.J. and M.C. Pinar. 2009. Pricing American Perpetual Warrants by Linear Programming. SIAM Review, Vol. 51, No. 4, pp. 767-782.

Wallace, S.W. 2010. Stochastic Programming and the Option of Doing It Differently. Annals of Operations Research, Vol. 177, No. 1, pp. 3-8.

Walls, M.R. 1995. Corporate Risk Tolerance and Capital Allocation: a Practical Approach to Implementing an Exploration Risk Policy. Journal of Petroleum Technology, Vol. 47, No. 4, pp. 307-311.

Walls, M.R. 1996. Developing an Exploration Decision Support System (EDSS): a Strategy for Combining Information and Analytics. Nonrenewable Resources, Vol. 5, No. 3, pp. 181-192.

Walls, M.R. 2004. Combining Decision Analysis and Portfolio Management to Improve Project Selection in the Exploration and Production Firm. Journal of Petroleum Science and Engineering, Vol. 44, pp. 55-65.

Walls, M.R. and J.S. Dyer. 1996. Risk Propensity and Firm Performance: a Study of the Petroleum Exploration Industry. Management Science, Vol. 42, No. pp.

Walsh, M.P. and L.W. Lake. 2003. A Generalized Approach to Primary Hydrocarbon Recovery. Elsevier Science, Amsterdam, The Netherlands.

Wang, S.S. 1998. Aggregation of Correlated Risk Portfolios: Models and Algorithms. Proceedings of the Casualty Actuarial Society Casualty Actuarial Society - Arlington, Virginia. LXXXV, pp. 848-939.

Wang, T. and R.D. Neufville. 2004. Building Real Options into Physical Systems with Stochastic Mixed-Integer Programming. The 8th Real Options Annual International Conference in Montreal. Canada, June 2004.

Winkler, R.L. and R.T. Clemen. 2004. Multiple Experts vs. Multiple Methods: Combining Correlation Assessments. Decision Analysis, Vol.1, No. 3, pp. 167-176.

Zenios, S.A. 2007. Practical Financial Optimization: Decision Making for Financial Engineers. Blackwell Publishing Ltd, Cambridge.

Vita

Wei Chen was born in Luliang County, Yunnan Province, China on June 16, 1974, the son of Bowen Chen and Guanzhen Zhu. Wei was married with Dr. Shuming Liu in 2009. Wei obtained his bachelor degree in the Department of Automobile Engineering at Tsinghua University, China, in 1998. Then he joined the School of Economics and Management at Tsinghua University as an instructor and undergraduate advisor. Wei received his master degree in Management Science at the same school in 2002. In the Fall of 2002, Wei entered the McCombs School of Business at the University of Texas at Austin to pursue his Ph.D. degree in Management Science. In 2008, Wei worked as a senior consultant in the Global Energy Group at Deloitte Touche.

Permanent address (email):

212A Millwood Dr., Millbrae, CA 94030 (wchen.sf@gmail.com)

This dissertation was typed by Wei Chen.